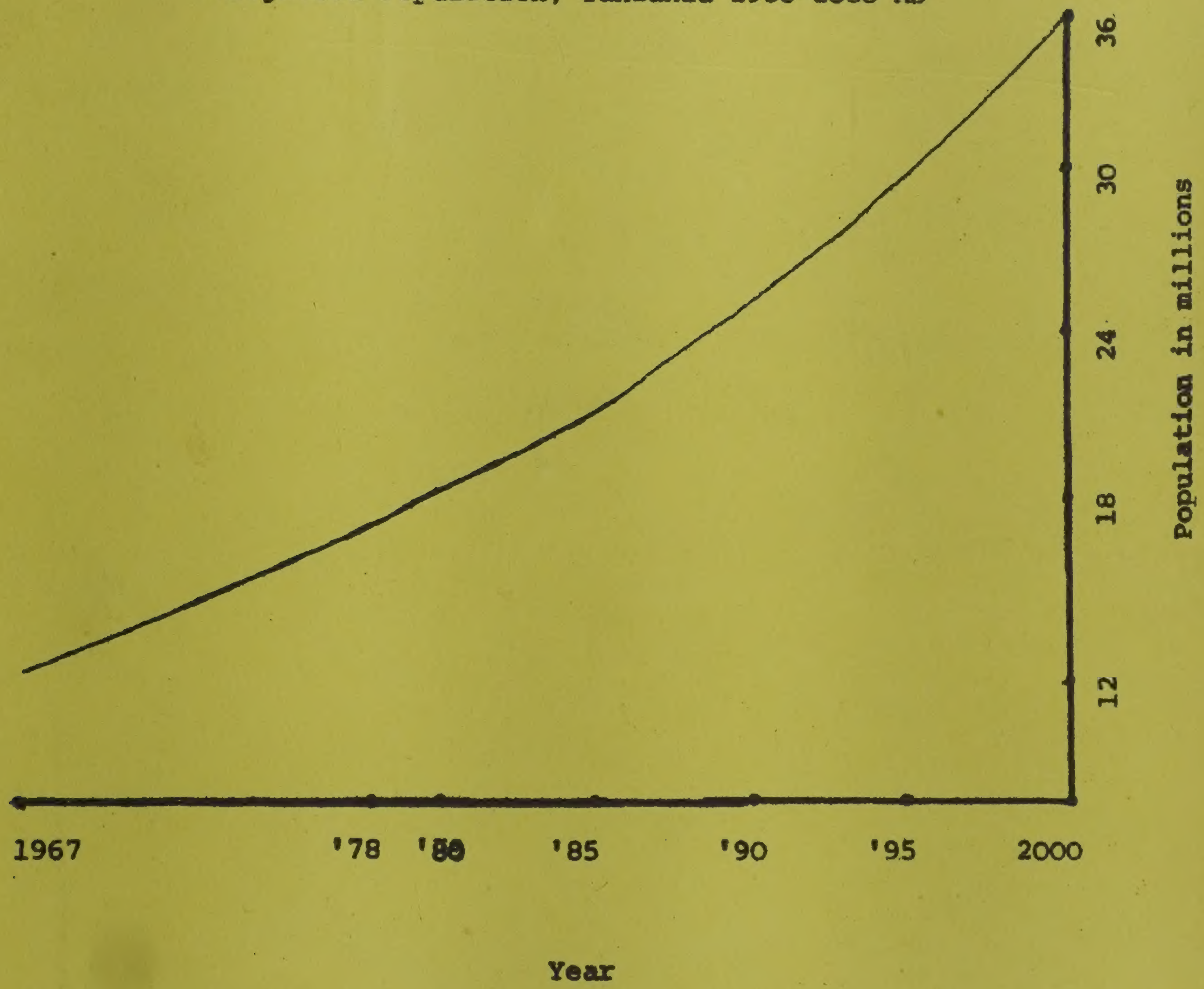
Mr. P. PADMANABHA

P. Padmanabha

WORKBOOK ON DEMOGRAPHIC ANALYSIS

Middle Level Course

Projected Population, Tanzania 1980-2000 AD



DEPARTMENT OF STATISTICS, ZANZIBAR

The United Republic of Tanzania

IN COLLABORATION WITH

ECONOMIC COMMISSION FOR AFRICA, ADDIS ABABA

F O R E W O R D

One of the serious problems faced with by the young (established in 1978) Statistical office in Zanzibar has been the shortage of trained manpower, especially at the middle level. This has affected its work both quantitatively and qualitatively. We have been caught in the viscious circle of low level of training of the manpower resulting in difficulties of getting them trained in the available regional training institutes which in turn results in their low level.

It was in this context that when the UNFPA Deputy Representative at Dar Es-Salaam and the Regional Adviser from the Population Division of ECA visited Zanzibar in August 1982 that a request for assistance for training of middle level statistical manpower in demographic and social statistics was broached. Statistical clerks working in the ministries of health, education, manpower and planning needed training to carry out their responsibilities more efficiently. Their response was quick and positive and the proposal took shape in the form of the training workshop for 6 weeks starting 5 April 1983 with 16 participants (5 from the Mainland and 11 from Zanzibar).

The course outline was prepared to reflect our needs and covers the various aspects of data needs, data collection, compilation, presentation, evaluation, analysis, interpretation and utilisation. The method of presentation of the materials was deliberately made practice oriented and illustrated methods particularly using Tanzanian data where ever possible. Discussion groups, seminars, field visits, etc., formed part of the programme in addition to formal lectures and laboratory exercise.

It was felt that the presentation of the edited version of the materials utilised in the training workshop will not only benefit those in Tanzania but also those in other countries faced with similar manpower situations.

We are thankful to the UNDP Dar-Es-Salaam and especially to Ms. Wilma Goppel, UNFPA deputy representative but for whose enthusiasm and interest this workshop would not have taken off the ground. To the UNECA, Addis Ababa and particularly to Dr. K.V. Ramachandran we are indebted for preparing the outline and time schedule of the course and taking on a lions share of the teaching programme. The Regional Institute for Population Studies, Ghana and especially Dr. A.F. Aryee was involved from the very early stages of the preparatory work of the programme and we are obliged for their cooperation and help.

A training programme of the nature we had organised needs the expertise of several persons and we are indeed lucky that we could draw on experts from the CSO in Dar-es-Salaam and Zanzibar, the University of Dar-es-Salaam, UNECA and other international agencies. Since they are many, it is difficult to mention them individually. A list of all the persons who assisted with the workshop is provided in the programme outline.

To the Ministry of Planning and particularly the Minister of State (Planning) Zanzibar and colleagues at CSO (Dar-es-Salaam and Zanzibar) we owe a debt of gratitude for cooperation and assistance. We are also thankful to the Director, Institute of Kiswahili and Foreign Languages for readily making available facilities for lectures /seminars/laboratory/assembly rooms etc.

It would have been impossible to carry out the workshop successfully without adequate number of calculating machines. We take this opportunity to thank the East African Statistical Training Centre and especially Mr. Muba, Acting Director for not only providing these machines but also taking active part in the training programme. The cooperation of Ministries in sponsoring candidates for the training for such a long period is very much appreciated. We hope that the time spent by their officers would result in wider experiences, more efficiency and their better utilisation in their work. To the participants we are thankful for their interest and active participation in the various activities which were crowded into the short period of 6 weeks.

Finally we take this opportunity to thank the United Nations Fund for Population Activities for funding the workshop and thus investing in our developmental activities.

Zanzibar
7 May 1983

ALLI ATHMANI
COURSE DIRECTOR

PREFACE

In this workbook are presented an edited and condensed version of all the materials given at the six week middle level inservice training workshop on demography and social statistics, Zanzibar-United Republic of Tanzania. It tries to cover the most important aspects of data collection, preparation, presentation, evaluation, analysis, interpretation and utilisation. In so far is feasible, technical details have been omitted-the focus being to make available to middle level statistical clerks in developing countries some of the methods which may enable them to analyse data.

Several persons participated in the workshop and the various sections were prepared by them as per an outline detailed in the next few pages. Professors Maro, Mlay, Namfua and Sembajwe from the University of Dar-es-Salaam; Mr. Muba from the East African Statistical Training Programme, Dar-es-Salaam; Messrs. Maimu, Mbaruku, Ngallaba and Uiso from the Central Bureau of Statistics, Dar-es-Salaam; Messrs. Athmani and Msellem from the Statistics Office in Zanzibar; Mr. Noah, Ministry of Education and Mr. Tenende, Ministry of Planning, Dar-es-Salaam; Mr. Blankert, Ministry of Health, Zanzibar; Dr. Aryee, Regional Institute for Population Studies, Accra-Ghana; and Messrs. Banda, Ekanem and Ramachandran, Population Division of Economic Commission for Africa, Addis Ababa-Ethiopia took part in the training programme.

Illustrations and examples are taken from the censuses of Tanzania especially the latest one of 1978. Since the holding of the workshop, the analytical report of the 1978 census of Tanzania has been published as Volume VIII of the census. A wealth of information is available in this publication and the reader interested in more details is referred to this important source.

This being a purely teaching material with no claim for originality, ideas and materials have been liberally borrowed from the existing text books, manuals, monographs and other publications. We acknowledge indebtedness to these sources.

Attempt has been made to cover the subjects in as thorough a fashion as necessary within the constraints of space and level of presentation. Since similar training programmes may be mounted in other parts of the continent, it is hoped that this publication would serve as a guide. Any comments, suggestions or criticisms would be gratefully accepted in order to improve similar presentations based on future experiences.

Addis Ababa
Dated : 9 November 1983

K.V. Ramachandran
Editor and Programme Coordinator

CONTENTS

1. INTRODUCTION

1.1. STATISTICAL SITUATION IN TANZANIA: NEED FOR DATA AND TRAINING - HISTORICAL PERSPECTIVE:

Organized Statistical services covering Tanganyika, Kenya and Uganda have existed since 19th July, 1926 when the Statistical Section of the East African Governors Conference was inaugurated to form the first statistical office. The unit was to centralize Statistical Organization and its function was said to be, "to collect statistics gradually, on the same method throughout the territories and to tabulate and compare the results so that true inferences can be drawn." A "Quarterly Bulletin" was first published in 1927 and continued for four years. After a suspension during the depression, on 1st January, 1948 an East African Statistical Department was established as a scheduled service of the East African High Commission its function being to institute general statistical services throughout the territories. It was to ensure comparability of data and similarity in timing and coverage of surveys. A territorial sub-office for Tanganyika was set up in Dar es Salaam in 1949 followed by a new section termed the Tanganyika Development section created in 1952 to be financed by Tanganyika Government. The first Statistical act can be traced back to April, 1949 when the Central Legislative Assembly gave powers to the Director of the East African Statistical Department to collect statistics on most economic subjects once the consents of the Administrator of the High Commission and the Central Legislative Assembly have been obtained. From 1st July 1956 the East African Statistical Department had five units, i.e. the three territorial units headed by Deputy Government Statisticians, the East African unit headed by the Deputy Director and the Statistical Adviser who directed the statistical organization and advised the East African Governments and the East African High Commission.

With the independence of Tanganyika, the East African Statistical Department split up and territorial Units taken away to form the nucleii of separate statistical offices of the three East African Governments. The statistics Act Cap 5 of the Laws of the High Commission was amended by Act No.2 of 1961 which transferred to the governments of the three territories many of the subjects listed in the original schedule. The East African Statistical Department remained covering specifically East African Statistics such as international trade (because of customs union), finance (because of East African currency Board), and the co-ordination of information on an East African basis (because of existence of East African Common Services Organization later to be the East African Community). The department died with the death of the East African Community in 1977.

The Statistical Ordinance of 1961 established the Statistical Service of Tanganyika from 1st July, 1961. The Statistics Division was renamed the Central Bureau of Statistics with a strength of 35 members:

headed by a Government Statistician. In Establishment Circular number 8 of 1967 the Government published its plan to centralize all government statistical services and to provide a scheme of service for statisticians to ensure a well staffed efficient service. By October 1968 the Central Bureau of Statistics had a staff of 114 employees divided into [illegible] sections corresponding to main branches of Statistics headed by Professional Statisticians. The principal statistical periodicals produced were the Annual Statistical Abstract and the Monthly Statistical Bulletin.

STATISTICAL ACTIVITIES:

(a) Demography

If we are to go by written records, population statistics seems to have the oldest record. The oldest population count or census recorded for Tanzania was in 1910 when African and Non-African population was counted in Zanzibar. The first record for Tanganyika was the population estimate made in 1913. Below is a table giving a summary of activities in this field:

Table 1. NATURE OF PAST ENUMERATIONS IN THE AREA OF UNITED REPUBLIC OF TANZANIA 1910-1978

Year of Enumeration	Tanganyika		Zanzibar	
	African	Non-African	African	Non-African
1910	-	-	Count (i)	Count (i)
1913	Estimate (ii)	Estimate (ii)	-	-
1921	Count	Census	-	Census (Apr.)
1924	-	-	Count (Mar-May)	-
1928	Count	-	-	-
1931	Count (Mar-June)	Census	Count (Mar)	Count (Mar)
1948	Census (Aug)	Census (Feb)	Census (Feb)	
1952	Partial Census (Feb)	Census (Feb)		
1957	Census (Aug)	Census (Feb)	-	
1958	-	-	Census (Mar)	
1967	Census (Aug.)			
1973	Demographic Survey			
1978	Census (Aug.)			

Sources Census Reports 1921-78

Notes: (i) includes Mafia Island

(ii) includes Rwanda and Burundi

.../3

For details of these activities the reader is referred to the original census reports. There seems to have been an achievement in being able to conduct a population count exercise once in a decade. Methodology applied changed with the times. The 1973 National Demographic Survey covering Tanzania Mainland only was directed partly at conceptual and methodological refinements in the context of data collection, provision of intercensal results and also studied the impact of mode of life on fertility and mortality.

In addition to the specific population census and survey reports, demographic data were also included in such periodicals as Monthly (later changed to Quarterly) Statistical Bulletins and annually in the Annual Statistical Abstract. The Annual Reports of the Registrar General gave data on births and deaths. An ad hoc survey on deaths, live births, still births and pregnancies was undertaken for all households in zones of Mkomazi, Valley, Mamba and Taveta/Kileo of Tanzania Mainland for the period 1962-6. Migration statistics was usually included in the Monthly Statistical Bulletins and the Annual Statistical Abstracts.

(b) <u>National Accounts</u>

In 1955 the "Tanganyika Gross Domestic Product" was published as a report of the Royal Commission on East Africa. The bibliography of the 1977-79 Statistical Abstract shows that this field has been well covered. The Statistical Service is well represented in the list of publications. Despite some gaps and time lag in publication, the period 1968-1980 is well covered. The field has the advantage of its data having to be included in the Background to the Budget/economic survey which has got to come out just before the budget every year; and has been up-to-date from 1955/56 financial year to-date. The appearance of data on national accounts in Quarterly Bulletins and Annual Statistical Abstracts depended on the publication of these two publications.

(c) <u>Trade Statistics</u>

Internal trade is one of the fields which seems to have been neglected by the Statistical Service. The only publication by the statistical service on this is the "survey of Distributive Trade, Dar es Salaam 1970". Otherwise to have information on the subject one has to resort to various administrative records such as local councils and revenue offices, co-operative movements, Marketing Boards, Crop Authorities, State Trading Corporation, Board of Internal Trade; and the Registrar General of Companies. These sources lack Statistical data collection methodology and analysis techniques.

On the International trade front the situation is a bit better. The East African Customs and Excise Department published Monthly and Annual Trade Reports covering interstate (between East African countries) and international (between East Africa and the Rest of the World) trade. The East African Statistical Department analysed these data and published them in the "East African Economic and Statistical Review". With the break up of the East African Community

in 1977, the Customs and Excise Department of the Treasury assumed the responsibility of the E.A. Customs and Excise Department. There is a problem of time lag in publication.

(d) Agricultural Statistics

For the collection of primary data, the Statistical service has conducted large scale commercial farming in 1960, 1962, and 1964; a sample survey in 1966 and the Agricultural Census as of 1950, 1960 and 1971/72. Otherwise most of the data appearing from the various publications are compilation of data collected mainly by Divisions of the Ministry of Agriculture. This Ministry has a Statistical Unit but the unit leaves a lot to be desired. There seems to be little co-ordination between Statistical Service and this unit. The statistical Service has little influence in the manner the data is collected or analysed. The Ministry produces a Monthly Statistical Bulletin of its own.

(e) Industrial Statistics

With some gaps, this field has been covered by the Statistical Service between 1958 and 1978 through the surveys of industrial production. Efforts to have annual surveys failed only during 1959-60, 1962-64, 1975-78 and 1979 to-date. However there were censuses of industrial production in 1958 and 1978.

(f) Labour Statistics

The coverage of this field by the statistical service spans the period 1961-1978; through the annual survey of employment and earnings. The Organization of Tanzania Workers, the only Trade Union, Annual Reports of the Labour Division and National Provident Fund are important administrative sources of data on this field.

(g) Social Statistics

The Statistical Service depends on the Statistical units of the Ministries of Health and Education for data on these areas. In addition to statistical publications by these Ministries, the service include the data in its periodicals the annual Statistical Abstract.

The Central Statistical Bureau however bears the responsibility of conducting Household Budget Surveys which also collect data in these subjects. Between 1956/57 and 1966 these surveys covered either certain urban areas or small chosen areas. It was in 1969 that a national survey was undertaken. The last one was undertaken in 1976 but its reports are yet to be published.

(h) Prices Statistics

"Prices and Price Index Numbers for Eighteen Towns in Tanzania" has been published annually from 1968/69; 1977/79; with gaps for 71/72, 74/75 and 76/77.

(i) Miscellaneous Statistics

In addition to the fields mentioned above, there have been other publications such as Report on Tourism Statistics in Tanzania for years 1968, 69 and 70 and a Report on airport Tourism survey in 1968 by the Bureau of Statistics.

(j) Periodicals

As indicated above, the Bureau of Statistics produces two periodicals, the monthly later changed to Quarterly Bulletin and the annual Statistical Abstract. After being published regularly in the sixties, the publication of the Abstract stopped in 1973. The latest publication covers 1973-79, which has just come out.

OBSERVATIONS

From the number of publications produced, it can be observed that the statistical service has been active. Differences in timeliness and number of publications between sections of the Statistical Service and between periods of time indicate that the service was more active in some sections than others and during some time periods than others. While in some fields the statistical service was active in collecting primary data, in others it relied on secondary data.

There should have been no harm in using secondary data if the service monitored or co-ordinated the data collection activities of the organs from which it got the secondary data. When there is no control, one may not be sure of the quality and comparability of the data, nor can one control the regularity of the availability of the data. The spirit of the Establishment Circular number 8 of 1967 of the Government seems to have died somewhere between 1967 and now. There seem to be no centralization of the statistical service. Where statistical units have been set-up, administrators have had a discretion of chosing to make them either independent from the service or to link it to the service. In some cases administrators have not known the alternatives available.

Amongst the problems facing the statistical service is the shortage of manpower. Of the 247 Statisticians and Statistical Officers required, only 73 are there. This has been partly because of the inadequacy of the terms and conditions of service of the Statistical Service. Not many people have opted to join the service. Also once in the service, the morale is low. Partly it is also because of the inadequacy of training programmes. For mainland, 332 need to be trained between 1981/82 and 1990/91. The output of the training institutions has not matched the demand of the Statistical Service. But given training opportunities, individuals have got to be attracted to take such training and ensuring that after training, the service they join is attractive for them to remain.

It would appear that in both the cases, government goodwill is the key to the success of the statistical service. Training institutions are

.../6

available, what remains is their utilisation. The Eastern African Statistical Training Centre, the University of Dar es Salaam and Makerere University are training institutions at the disposal of the government. Slight adjustments in Government priorities could enable the Bureau to have the needed equipment to keep it running and produce the essential data required.

CONCLUSION

Since the Statistical Ordinance 1961 entrusts the responsibility of heading the statistical service to the Government Statistician, his observation in 1980/81 of the situation of the service would make the best summary. "At the present time the Central Bureau of Statistics does not have the capability to collect and compile all the needed statistical information. What more, even the processing capabilities of the collected material are strongly restricted and this results in the wide time gap between the moment of collection of the data and the moment of their publication. The resources (manpower and equipment) available are not adequate. This situation, incidentally, is not new, in fact things have been this way for a number of years. The level of statistical production has fallen over the years since the early seventies and this have given rise to suggestions of establishing and strengthening the statistical units in ministries and other public institutions. However, we feel that the role of the Bureau should be strengthened in order to widen the scope of its own statistical production and to enable it to co-ordinate and supervise efficiently the statistical work of other institutions".

THE STATISTICAL SITUATION IN TANZANIA MAINLAND

The Need for Statistical Information

In Tanzania, just like in any other country, planning and policy making at government and corporate level are based on statistical as well as other information, the lack of which can render such an important task almost impossible. The need for Statistical information for the planning and management of the national economy manifest itself in the heavy demand placed on the Central Bureau of Statistics (CBS).

In the past, planning and decision making in Tanzania have been based on scanty and sometimes inaccurate statistical information, as a result of which planning has not taken all the relevant factors into account to make our national plans as comprehensive as they should otherwise be. Moreover, due to the past poor statistical base, the country's plan implementation machinery has not been as satisfactorily sound as to allow for speedy plan monitoring and evaluation.

With the current unprecedented economic problems facing Tanzania, the need for statistical information to facilitate well informed decision making at this crucial period in our post-independence economic history is not only the more important and obvious but also extremely fundamental to the whole planning process.

.../7

It is on this background that this paper examines, albeit briefly, the statistical situation in Tanzania Mainland. First, we look at the statistical organization with the Central Bureau of Statistics at the centre. Secondly, the paper looks at the statistical legislation that provides the legal framework for the collection, compilation, analysis and publication of statistical information and related matters. Thirdly, a review of past statistical performance is given to see how best or poorly the CBS has been meeting the challenge of statistical production to satisfy the growing demand for various statistics. The paper also will take a look at the present position to see what improvements and/or innovations are being made. In the last part we examine future plans that are intended to develop the various statistical series as well as the infrastructure.

STATISTICAL ORGANIZATION

The Statistical service in Tanzania Mainland consists of the Central Bureau of Statistics at the centre and several statistical units attached to ministries and other public institutions. Among these units - all of which deal with sectoral statistics - are: the Research and Statistics of the Bank of Tanzania, the Statistical Units in the ministries of Agriculture, Natural Resources and Tourism, National Education and Health, and the Statistics Unit in the Customs and Excise Department. Statistical data are also collected by institutions other than those mentioned above as part of their general administrative functions. However, such data are, to a large extent, neither tabulated nor published.

The main duty of the Central Bureau of Statistics is to collect, compile, and publish statistical and related information relating to economic, social and cultural activities and conditions of the people. It also collaborates with other public institutions in collecting, compiling and publishing statistical information.

The Central Bureau of Statistics consists of the head office and twenty regional offices. It is headed by the Government Statistician assisted by eight Assistant Government Statisticians. At the head office statistical work is organized under eight sections, each of which is headed by an Assistant Government Statistician. These sections are: (i) Population and Tourism Statistics; (ii) Labour and Price Statistics; (iii) National Accounts Statistics; (iv) Agricultural Statistics; (v) Industrial Statistics; (vi) Trade and Transport Statistics; (vii) Sample Surveys; (viii) Statistical Services. These sections with the exception of the last two deal with the collection of statistical data in their respective areas. The main duty of the Regional Statistical Offices is the collection of statistical data which are eventually sent to the head office for analysis.

STATISTICAL LEGISLATION

The legal framework for the collection, compilation, analysis and publication of statistical information and related matters is provided

for under the Statistics (Ordinance No.33) of 1961, and is amended from time to time.

Implicitly in the Statistics Ordinance, and in practice, the statistical service in Tanzania Mainland is a centralized one with the CBS controlling all statistical activities. The major advantage of this system is the better use of limited statistical resources. It is important to add that it is a centralized statistical service with some flexibility in the sense that other government ministries/departments and other organizations continue collecting and analysing certain statistics under their respective areas of competence. Thus, the Central Bureau of Statistics has the added responsibility of co-ordinating all the statistical activities undertaken by those statistical agencies/units. Therefore the effectiveness of the CBS derives from the Statistics Ordinance of 1961 under which it still operates.

REVIEW OF PAST STATISTICAL PERFORMANCE

With increasing economic and social activity, the inter-relationships between various sectors of the national economy become more and more complex, and it becomes necessary for policy makers to have well analysed and presented statistical information at their disposal in order to visualize the implications of various policy options. In an attempt to meet this challenge, the CBS has over the past two decades (the 1960s and 1970s), endeavoured to sustain, improve and expand the statistical base and statistical information system in general with a view to meeting the country's short-, medium- and long-term socio-economic planning requirements.

During this period the major sources of statistics have been the following surveys undertaken by the Central Bureau of Statistics:
(i) The 1967 Population Census; (ii) The 1969 Household Budget Survey (HBS); (iii) Agricultural Census of Tanzania, 1971/72; (iv) National Demographic Survey of Tanzania - 1973, (v) The 1976/77 Household Budget Survey; (vi) The 1978 Population Census, (vii) The 1979 Industrial Census. Various small scale surveys limited in scope, coverage and purpose have also been undertaken during the period.

Despite many setbacks and problems encountered during the past two decades, the CBS has nevertheless managed its operations with some degree of success.

The 1967 Population Census was published in six volumes:

Volume 1 - Statistics for Enumeration Areas,
Volume 2 - Statistics for Urban Areas,
Volume 3 - Demographic Statistics,
Volume 4 - Economic Statistics,
Volume 5 - Census Methodology,
Volume 6 - The Population of Tanzania: An Analysis

.../9

The 1969 Household Budget Survey was published in three volumes: Volume 1 - Income and Consumption; Volume 2 - Housing conditions and Volume 3 - Retail Prices.

The Agricultural Census of Tanzania - 1971/72 was one of the most problematic of surveys ever undertaken, so much so that the results were published almost after ten years. They are in three volumes: Volume 1 - Peasant Farming; Volume 2 - Large Scale Farming and Volume 3 - Summary of Peasant and Large Scale Farming.

The 1973 National Demographic Survey of Tanzania was published in six volumes: Volume 1 - Regional and National Data; Volume 2 - Data for Socio-Economic Groups; Volume 3 - Summary Data for Survey Cluster; Volume 4 - The Methods Report; Volume 5 - Training Manual and Volume 6 - An Analysis of the 1973 National Demographic Survey of Tanzania. However, since then no other demographic survey has been carried out.

To date no volume has been published on the 1976/77 Household Budget Survey. Nevertheless, the CBS has issued a paper on the Listing of Households for the 1976/77 Household Budget Survey relating to Cash Income and Household Size mainly for internal use. This state of affairs reflects vividly more than anything else the poor data processing capacity of the Central Bureau of Statistics.

As regards the 1978 Population Census, all the data processing has been completed. So far census volumes that have been published are:

Volume I - Methodology Report,
Volume II - Population by Age and Sex for Villages/Wards and Urban Areas,
Volume III - Census Geographic Work,
Volume IV - A Summary of Selected Statistics,
Volume V - Fertility and Mortality Data for Rural and Urban Areas of Regions,
Volume VI - Private Households and Housing Characteristics
Volume VII - Basic Demographic and Socio-Economic characteristics.

The last Volume VIII on "Population of Tanzania: 1978" was published in September 1983.

The processing of the 1978 Industrial Census is still in progress. Until now two volumes on the census have been issued. These are: Industrial Census (1978) Volume 1 - Directory of Industries, 1979, 10+ member Industrial Establishments and Volume 2 - Directory of Industries 1979, 5 - 9 member Establishments.

Thus, the Central Bureau of Statistics has been fairly successful in its operations during the past two decades, given the limited resources (manpower, equipments etc.) at its disposal.

PRESENT STATISTICAL POSITION

At the present time the Central Bureau of Statistics does not have the capability to collect and compile all the needed statistical information. The resources available are not adequate. This situation

.../10

incidentally is not new. In fact things have been this way for a number of years and particularly so during the seventies when the level of statistical production fell considerably.

However, since late 1980 the CBS have been given a new team of managers consisting of the Government Statistician and Assistant Government Statisticians. With the new leadership, the major pre-occupation of the CBS since then has been to update all the major publications issued by it. The current situation is very encouraging. Most of the major publications including the Statistical Abstract, 1973-1979 have been issued and others are due for publication anytime from now.

THE FUTURE

We are aware that the need for statistical information does not go hand in hand with the level of a country's development and that a developing country must use relatively more funds in data collection and analysis than a developed country. But the cost involved is unavoidable since decision making is not easy or even possible without adequate and relevant statistical information. Planning without adequate and relevant statistics is just like guiding a ship without the benefit of a compass; and the consequences are not much different.

From this awareness, the Central Bureau of Statistics has worked out a long-term programme to build up the statistical infrastructure and improve the various statistical series. The major objective is to ensure that the statis ics produced by the Bureau are adequate, consistent, timely and relevant.

The ten year programme is in two parts: - the first deals with problems related to the infrastructure of the CBS and the second part deals with problems connected to the statistical series (adequacy, timeliness, consistency, etc.). These issues are very much related: for improvements in the infrastructure in most cases lead to improvements in the statistical product.

In building up the infrastructure the following are considered in terms of members and probable costs:

(i) Manpower requirements and training
(ii) Equipments which include:

(a) transport equipment
(b) data processing equipment including calculators
(c) printing and related equipment
(d) weighing scales
(e) camping equipments

The accommodation problem is being solved, at least temporarily by making extensions to the CBS headquarters building.

CONCLUSION

It is clear that if the statistical programme proposed by the Central Bureau of Statistics is satisfactorily implemented, then a

bright new chapter will have been written in the history of Statistical development in Tanzania Mainland. Correspondingly, the CBS will be rightly looked upon to live up to its expectations in satisfying the growing demand for statistics.

THE STATISTICAL SITUATION IN ZANZIBAR

As we all know for any sound planning and policy formulation the role played by Statistics is extremely important. In the case of Zanzibar the great demand for Statistical information became evident during the preparation of the first post independence three years Development plan which took off in 1978. The present Statistical Department was therefore established in July 1978 with the following points of reference.

(i) To compiile, analyse and publish all economic and social statistics.

(ii) To co-ordinate the statistical services which are being carried out by the statistical agencies like - Education, Health, Agriculture, Finance, Natural resources, department of migration, customs, income tax, etc.

(iii) To educate all those who are concerned on the importance of and use of Statistics.

(iv) To introduce a legislation which will empower the Department to collect the various statistics required.

The Statistical Department · Zanzibar although it is only four year old has managed to meet the challenge for which it was formed and in doing so it has been able to bring out some statistical series as well as the development of the institutional structure.

The Organization of the Department

The Department of Statistics - Zanzibar performs the central statistical service role but co-ordinates with some Statistical units which are attached to the ministries, departments and parastatals.

The Head of the Statistical Department is the Director of Statistics who is assisted by three Heads of Sections who are senior stasticians. The three sections are:

(i) Statistical Development Section which is charged with the carrying out of censuses and surveys.

(ii) Statistical Production section - this is the largest section of the three and it consists of 6 units namely:

(a) Agricultural statistics unit; (b) Industrial statistics; (c) Trade and Transport statistics unit; (d) Labour and Employment and Price statistics unit; (e) Population unit; and (f) National Accounts unit.

(iii) Statistical Services Section - consists of two Units which are Documentation and Information and Data Processing.

There is also a Statistical Office in Pemba whose responsibility is to collect all the statistics pertaining to Pemba Island and send them to Head Office Zanzibar.

STATISTICAL LEGISLATION

The Statistics are collected nunder the Zanzibar Revolutionary Council Decree No.7 of 1979. This decree is based on the amendment of the statistical ordinance No.33 of 1961 used in the Mainland so that it can also be applied to Zanzibar.

WORK DONE SO FAR

The Statistical Department has managed to carry out three main Surveys/Census during this period. These are: The Population Census which was carried out in collaboration with the CBS in 1978. The processing of these data has been carried out on the Mainland and the number of publications already completed have been included in the statistical situation of mainland.

(i) The Industrial Census - also this has been taken in collaboration with CBS and the reference period was 1979. It covered establishments engaging 5 or more people.

(ii) The Household Budget Survey 1982 is the first such survey to be carried in the Islands. Data collection has been completed and the analysis for housing conditions is ready except for income and consumption which will have to be processed at the CBS through the computer.

However, there are other surveys being carried out regularly. The Employment and Earning Survey is taken annually. No publication has been compiled yet on this survey but some of the data on this subject can be found in the Statistical Abstract, Zanzibar. The Annual Survey of Industrial Production has been carried out since 1978 and published upto 1981. The other survey concerns the public enterprises, the statistics of which are published in the "Analysis of Public Enterprices". So far two such publications have been printed; one for 1978 and the other 1979-81. Besides the above survey publications the department has brought out the statistical Abstract Volumes I and II, and the National Accounts of Zanzibar.

FUTURE

Although there have been some achievements, the department is still faced with a number of problems which will have to be solved for the efficient running.

As explained earlier the department is only four years old probably the youngest in Africa - so it is faced with acute trained manpower practically in each field of statistics. To solve this there is a statistical training programme for the personnel.

.../13

At the moment there are no computer facilities within Zanzibar so it is recommended at least to have a small type of computer (desk) to start with which could handle the small volume of statistical data.

Like any other statistical department there is the problem of transport equipment and other equipments (like calculators etc.). This will take time to solve as the country is faced with worst economic conditions.

On the shortage of office accommodation which has prevailed for sometime now the Ministry of State (Planning) which is the parent Ministry of the department is on an expansion programme of the office building and it is hoped this will help to some extent in solving this problem.

In conclusion I would like to say that with those existing constraints the department have lived to the expectations for which it was created. At least there are some statistics;what remains is to improve them so that they can help the planners in their efforts to draw sound and reliable plans which will facilitate well informed decision makin .

1.2 PURPOSE, HISTORY AND SCOPE OF DEMOGRAPHIC AND SOCIO-ECONOMIC DATA

The history of the development of the systematic collection, study and analysis of population data as a distinct field of study can conveniently be divided into four main periods. There is firstly, the period dating back from recorded history up to the time of the Englishman John Graunt (1662). There is secondly, the period from the time of John Graunt up to the advant of regular census taking (around 1800). This period also produced Malthus whose writings on population made a profound impact on the study of population phenomena. There is thirdly, the period of census taking (i.e. the early part of this century (1920-25)), and lastly the modern era when the most significant developments in population analysis have taken place.

Each of these periods marks a distinctive stage in the development of our state of knowledge concerning population and its interrelationship with other phenomena.

The first period is distinguished by the absence of any systematic attempts to study population phenomena. Various philosophers, historians, geographers etc. did write or commenton various aspects of population phenomena. Writers such as Plato, Aristotle, Ibn Khaldoun, Confucius all made some very perceptive comments about population and its interrelationship or implications for political organization, food supply, social life etc. But these writings were very largely speculative, and lacked many empirical or scientific basis. It is not surprising therefore that until the 17th century, there was no word in any European language for population as a concept. It was not until 1612 that an English philosopher first used the term as a distinct concept, and it was not till 1855 that the term 'demography' was coined by the Belgian, A Guillard in his book "Elements of human or population statistics".

John Graunt inaugurated a whole new era in the scientific study of population phenomena with the publication of his book, the

.../14

"Natural and Political Observations Made Upon the Bills of Mortality" in 1662. Graunt's important work was followed by several others such as William Petty, Gregory King and Sussmilch. Malthus' important essay on population also came out in 1799.

Though some censuses were carried out in many countries before the end of the seventeenth century, regular or periodic censuses really started at the end of the 18th century. They began in 1790 in the United States, and in 1801 in both France and England. Thereafter census taking regularly spread to all parts of Europe and even the colonies.

The more systematic study of population phenomena using various mathematical tools or models started with two important publications in the 1920s. The first was the publication of Carr-Saunder's "The Population Problem - A Study in Evolution" in 1922, and the second was the publication of Lotka's "On the true rate of natural increase as exemplified by the United States" in 1925. The first book argued strongly for an interdisciplinary approach to the study of population phenomena and hence marked the beginningof the social science approach to the study of population. The second on the other hand, demonstrated the need for the development of statistical models, or techniques for the study of population, and therefore both in one sense provide the foundation on which modern population study is based. The interest generated by all these developments led to the First World Population Conference held in Geneva in 1927, followed by the founding of the IUSSP in 1928.

SCOPE OF DEMOGRAPHIC DATA

Our primary aim in the collection of demographic data is to answer three basic questions. The first question relates to the size of the population itself, whether it is growing, declining, stable and the causes, significance, implications or consequences for this size or rate of growth.

The second question relates to the varieties and categories of people found in the population and how these differ from those in other populations. These characteristics relate to differences in age, sex, marital status, economic activity etc. The third major question relates to the distribution of the population or where people are located, and the reasons, implications or consequences for such distribution.

The questions listed above can be answered by looking at five major demographic processes taking place within the population. These are fertility, mortality, migration, nuptiality and social mobility.

To understand these demographic processes adequately, it is also necessary to study the social characteristics of the population such as education, literacy, ethnicity, religion, housing conditions.

1.3 USES AND IMPORTANCE OF DEMOGRAPHIC DATA

The uses and applications of demographic and socio-economic data vary with the needs of the various users - government, business, public and private organizations, the society, the community and even the individual citizen.

By far, government is the largest consumer of these data and one of the most important use the government has for the information is for planning. In developing countries with limited time and resources (men and materials) it is important that there be a rational utilization of available resources. Since planning involves evaluation of past and current situation, performance, demand, utilization, attainment and change etc and aprojection of these into the future the larger the amount of information available and the better its quantity, the more appropriate the planning can be.

The basic data needed are population size, structure, composition, location, distribution, dynamics and other characteristics because population is not only the producer and consumer of goods and services and reproduces future generations, it is central in the planning process.

These information are also useful for the business men and others to plan for their activities.

Another important function of the data is to enable the formulation of policies which will accomplish the goals of national development as envisaged in the plans by identifying current and future needs of various sectors of the population for services, opportunities, amenities and facilities, to establish priorities and take action programmes and evaluate the progress of the programme.

Demographic and socio-economic data also are useful in the carrying out of administration and implementing policies and plans.

Population is one of the most important resource of a country. Its development and fuller utilization benefits the nation, the society and the individual.

The size of a population indicates the magnitude of the human resources and determines population density and the relation between available resources and potentiality for production and consumption of goods and services and is a determinant of the reproductive capacity.

Since production, consumption and reproduction and a whole host of attitudes, behaviour and practices are determined by age and sex of the individual, it is essential to have information on population structure also in addition to size. Several socio-economic characteristics like education, economic activity, marital status, family formation, etc impinge on the individual and the society and have important ramifications on the quality of the population. Hence, it is imperative that in addition to age and sex and size of the population, we also have information on population composition in respect of characteristics which have relevance in particular contests.

The geographic location of a population could be the cause or consequence of several factors not the least important of which could be climate, vegetation and man made artifacts like location of industries, educational, recreational and other facilities and so on. Allied to concentration of population is the problem of population dispersal. For effective planning and policy making it is essential to have information on the spatial aspect of a population.

Populations grow or decline through the dynamic of births, deaths and migration. These determine and are determined by the other characteristics of the population like the age - sex, marital status, level of education, economic activity, spatial distribution and so on.

Thus we note that a study of population is complete only when not only the size but the quality, distribution and change are also included and all these information are needed for planners and policy makers to estimate the types and quantum of social, economic and other programmes needed. For example, the provision of health services require information on the size of the population and their location. In addition, since the needs vary from one age to another and differ by sex, we also need the age - sex distribution. One of the important factors in health service need and utilization is the level of morbidity. For example, the number of hospital beds, doctors and other health aids needed would be determined by the incidence and types of diseases and disabilities in the population. The demand and utilization may, on the other hand, be determined by socio-economic, geographic and psychological factors.

Again provision of education is determined by age - sex distribution of a population, the attitudes, aspirations and participations which in turn are effected by socio-economic and spatial aspects. The types of education will be determined by the socio cultural environment and the plans for development.

This is true of economic activity and participation in the prodction and consumption of goods and services.

As a matter of fact, demographic and socio-economic data are central requirements in any planning and policy formulation exercise and with limited time and resources become very important for rational decision making.

2. DATA COLLECTION

2.1 SOURCES OF STATISTICAL DATA

Introduction

The raw materials for a statistical activity are the statistical data. Amongst the functions of a statistician is to procure the raw materials to be able to produce statistics. The exercise of data collection takes a statistician to various fields of study ranging from pure science, social sciences and other fields of study. They are all related in that they affect the population.

Types of data

Statistical data can be classified into three major categories. Economic statistics which cover the economic aspects of peoples life, and include basically statistics on production, distribution, consumption, incomes employment and prices. On the other hand, Social statistics cover social aspects of the conditions of life and work of populations, and are concerned with the status and changes in status of various population groups in regard to such factors as occupation, earnings, housing, education, health, recreation, family life, community activities, cultural activities and interests and by extension of social life, justice and crime. Geographical or spatial aspects are yet another category.

Data Collection Agents

(a) Government. We have seen that statistical data cover various aspects of life. For a long time now, people living in societies have entrusted the responsibility of taking care of the interests of the society to systems of organizations called governments. Governments are supposed to ensure the internal and external security, national economic and social development, public welfare and social justice. Government has been given not only the responsibility but also the power and the means to do these. Since Government would need statistics to facilitate its administrative operations and to monitor the efficiency of those operations as well as to evaluate and reformulate its policies, it is only right to expect the Government to collect the needed statistics.

A good amount of statistics would arise as a by-product of administrative operations such as: Registration of births, deaths, marriages and divorces leading to vital statistics. Registration or licensing of business undertakings produce basic frame of economic institutions. Control and regulation of the functioning of such economic institutions lead to the accumulation of pertinent operational statistics, which selectively produce related national aggregates. Direct developmental activities in transport and communication, health, education and other welfare measures lead to statistics in these fields. Most Governmental activities are associated with relevant revenue earning measures and they intend to ensure good coverage. But sometimes they omit non-governmental sections. Statistics collected by Government as part of their administrative activities, are often fairly available in concerned Ministerial or Departmental Reports.

A government statistical service is established to ensure that data collection fulfil the basic statistical principles. It would supervise and co-ordinate the data collection activities of other government institutions to ensure that data produced meet acceptable standards in terms of coverage, comparability and quality. A synoptic picture of usually the whole statistical information (together with indications of primary sources) is included in the Annual Statistical abstract. This serves as a basic reference for information over time and indicates broadly the nature, classification and presentation of available information.

Government Statistical service or other government departments conduct, from time to time, according to need and usage, special statistical enquiries. Decennial Population Censuses, Periodic Censuses of Manufacturers, Surveys of Farm Production, Employment and Earning Surveys and Household Budget Surveys are examples of such enquiries. Results are published in related reports.

(b) Business and Other Economic Organizations. For efficient management, most economic organizations - public and private - need to keep statistical records. These records may pertain to production, sales, supplies, personnel, financing of undertakings and costs of products. Occasionaly, large scale undertakings make special studies regarding consumer opinions, sales patterns, market demand, plant utilization, labour supply, labour absenteeism and so on. These records may or may not be published in appropriate summary forms but they constitute an important basic source of economic statistics. These summary information arising from these sources become available in large company's Annual Reports. In any case the salient parts of this information flows into the Government in fulfilment of Government controlling and/or regulatory enactments.

(c) Research Studies or Investigations. National or international individuals or bodies carry out research studies to give additional information or interpretation from new angles. Such bodies as Universities and the National Scientific Research Council are examples of this group. In developing countries this group is not very large.

Data Collection Methods

(a) Census. Statistically, primary data may be collected by observing all units that qualify to be included in the population. This method aims at having complete information free from sampling errors but costs involved and degree of non-sampling errors which are difficult to evaluate make such enormous exercises unnecessary.

(b) Sample Surveys. Statistical methods have shown that information about a population can be obtained by observing only part of the population instead of the whole. This reduces non-sampling errors though it introduces sampling errors. Sampling errors can be determined and precisions and reliability of results improved. One has a choice among many sampling schemes to suit the circumstances.

(c) Registration. As an administrative process, recording of events as they take place is one way of collecting information. Such activities as registration of births, deaths, marriages, companies etc. is a good method of data collection, if good coverage is achieved.

(d) Other Administrative Records. Administrative records maintained by the administrations of organizations for their own day to day operations such as accounting records, operational activities and so on can be a good source of statistical data.

Conclusion

Data collection is the first and most important process in the production of statistics. We have seen that various agents collect data for various objectives - statistical and non-statistical. We have also seen that various methods of data collection are used. The quality and reliability of statistical data produced would depend on the way the agent collected the data and on the objectives of collecting the data.

2.2 DEFINITION OF A POPULATION

The main objective of any census is to count and record the personal characteristics of all the persons living within its territorial borders at a specified time. How successful a census is therefore depends on the following considerations: firstly that all persons within the territorial boundaries are accurately accounted for, and secondly that some persons are not counted more than once for one reason or the other.

Censuses are undertaken by governments mainly for planning purposes, and effective planning is done not only for the aggregate or total population but more importantly for its sub-units such as districts, regions and localities and for segments like infants, children, women, old persons, etc.

The method used in enumerating or counting or "defining" the population determines to a large extent how accurate the information is and also how far the data collected can be used for the type of regional and segmental planning mentioned above. The two main methods of defining the population to be counted are the 'De facto' method and the 'De jure' method.

De-facto method: The 'de facto' method aims to count and record particulars of all persons who were in the dwelling, house, or area at a specified time, e.g. census night. This method, sometimes referred to as the "place where found at time of census" or "present-in-area" method, implies that the census is taken with reference to a particular time or period usually referred to as census night. This means that even if the census lasts for a period of two weeks, the information collected on all persons should relate only to the areas, or houses or dwelling units in which they spent the census night.

'De-jure' Method: The de-jure method of counting on the other hand aims to count and record particulars of persons at the places (area, houses or dwelling units) where they usually live, or places of "usual residence". This implies that the census enumerator is expected to exclude from list all persons who are present in the area or house at the time of his visit but are merely visiting. By the same token, he will count all persons who have travelled or are away at the time of his visit, but who normally reside in the area.

.../20

Evaluation of the De-facto Method

It is generally agreed that the main merit of the de-facto method lies in its simplicity and ease of operation. Theoretically, every person should be able to remember where he was on census night - the reference period - so through such simple devices as asking the head of the household to note persons who spent the reference night in their houses, it is possible to count every person in relation to only one specific area and thereby avoid duplication or omission.

The work of both the interviewer and the respondent is simplified because all they have to ask or remember is where the respondent was on census night. Thus if the period of enumeration is kept very short, it is reasonable to suppose that remembering this would not constitute much of a problem especially if the census night had been adequately publicised - such as by bonfires, tolling of church bells, special radio and television programmes etc. - to make a lasting impression on the minds of the populace.

The use of the de-facto method however presents some problems. The first problem relates to the various categories of people who are not at precise locations or places on the census night. Even though, through legislation and other measures, a large proportion of the population can be induced to stay indoors on census night, there will still be, on census night, a number of people out fishing, hunting, on the road, in the air etc.

There is also a second category of persons who are at precise locations or places, but these are not normally places the interviewer is expected to call at during his usual rounds; examples of such persons are those at an all-night dance, those who sleep in the markets, under bridges, in hospital etc.

For de-facto enumeration to be successful, special techniques must be adopted to enumerate all such persons within the shortest possible time. The techniques vary from group to group. For those at a dance for example, it is possible to relax the strict rule about the "place-where-found" on census night and enumerate him at his usual place of residence, while for the "floating" population who sleep at markets, under bridges etc., a special intensive exercise to enumerate all of them in one night is undertaken.

The second major problem in the use of the de-facto method is the result of the phenomenon generally called "census migration". In some African countries where censuses are of recent origin, some chiefs and politicians often see censuses as an opportunity to procure more amenities and services for their areas by inflating their normal populations. They may therefore persuade their citizens residing in other areas to return home for the census.

As may well be imagined, if this happens on a very large scale, it may completely distort the geographical distribution of the population,

It may also lead to serious undercount or overcount because some of the people (including whole families) may leave before the census is over, and may therefore be missed both in their places of origin and places of present residence. Of course, the reverse may also occur.

Another disadvantage of the de-facto method, if it is strictly carried out, is that it does not always give a true picture of the permanent population. For example, in countries such as Botswana and Lesotho where a considerable number of their citizens live and work in neighbouring countries such as South Africa, a strict application of the de-facto principle would mean that no information would be collected on such persons, although they form part of the population. This also means that if there is a large population of temporary residents - like people attending a conference, an army providing temporary military assistance - they would also be counted as part of the population.

Though the numbers involved may not be large, this could still lead to serious problems. For example, the doctor population of the country could be inflated by two hundred if an international conference of doctors was taking place at the time of the census, thereby giving a very wrong impression of the state of medical facilities in the country.

Another disadvantage of the de-facto method is that in order to complete the enumeration as quickly as possible, a large number of interviewers may be required thereby increasing the size and cost of the census machinery, and even in some cases jeopardizing the quality of the information collected.

In spite of these disadvantages, the de-facto method remains the more popular method because of its simplicity.

Evaluation of the De-jure Method

Since the de-jure method counts people at the places where they normally reside, it generally gives a more accurate spatial distribution of the population, and is therefore more useful for planning at the local level.

There is also no need for respondents to remember where they were at a particular time, even though the information collected may still relate to a specific reference period.

The use of de-jure method also obviates the need to complete the enumeration in the shortest possible time as in the case of the de-facto method; this means that a smaller but more qualified number of interviewers may be used, thereby improving the quality of the information collected.

The use of the concept however raises some problems which are particularly acute in African societies owing to the nature of our

social life. The first problem relates to the meaning of the term "usual residence".

In a society in which people regularly commute between the locality of origin and the locality of settlement, between the fathers' compound (which may be at different locations), between the husband's house and the parental home etc., it is not really easy to determine with any degree of precision what the usual residence is, and in many of these cases, the respondent may be regarded as "usual resident" in both dwelling units.

The second problem relates to the period of residence in present locality. In other words, how long must a person have lived at present locality in order to regard him as a usual resident? Some may have moved only a month ago on transfer into their present dwelling units and intend to live permanently in their new localities. To aoid the possibility of such persons being regarded as "usual residents" at both their old and new places of abode, a fixed period of continuous residence, usually six months, is specified. However, the period chosen for inclusion/exclusion of members of households should be same.

This of course does not completely solve the problem because interviewers do not sometimes have the patience to ask the elaborate and detailed questions which the use of this concept entails.

Another major problem with this method is that a certain proportion of the population may not live in "residences" at all. In some of the major cities in Africa, a large number of migrants do not have any regular places of abode and sleep at railway stations, lorry parks, markets etc. Unless special measures are taken to enumerate such persons, a substantial number of them may be missed in the de-jure method.

Modified method

As the above discussion has shown, both methods have their merits and demerits, and the more appropriate method to use will depend on the conditions in a particular country. The disadvantages of the two methods can however be modied by combining both methods in the same survey or census. While the one section of the questionnaire can list all those who spent census night in the dwelling (including visitors), a second part of the questionnaire will also list all those who usually live in the dwelling but were away on census night.

The use of this comprehensive enumeration technique makes it possible to obtain information on both the de-facto population and the de-jure population. Thus, although it involves more work, it tends to minimise the possibility of coverage errors.

METHODS OF ENUMERATION

There are two major methods of collecting the desired information in any census or survey; these are the Direct Interview or Canvasser Method, and the Householder or Self-Enumeration Method.

In the Direct Interviewer method a trained interviewer representing the organization or office conducting the survey visits the respondents in their own houses and then asks, and records the required information on a questionnaire. Though it is usually desirable in such cases to interview all the respondents individually and personally, it is often impossible to do this for various reasons.

Some household members may have gone to work, some women may be in purdah and cannot therefore speak to the interviewer, and others in the household may be minors; for all these reasons, even in the direct interviewer method, the information for the whole household may be supplied by just one or two persons.

In the householder method on the other hand, the questionnaires are delivered to the various households, and using the accompanying instructions, the head of the household is expected to record the desired information on the questionnaire in respect of all members of his household. The questionnaire is later checked and collected by an agent of the Census Office.

Evaluation of the methods. The major advantage of the Canvasser method is that it is possible to ensure uniformity and greater accuracy in the interpretation and recording of the required information. Interviewers are usually intensively trained or should be trained before they undertake any exercise. In the householder method, on the other hand, the householder may choose to ignore the accompanying instructions or fail to understand them properly, especially if some of the questions are complex. Furthermore, the trained interviewer has a greater commitment to the success of the exercise than the householder, who quite often, may regard it as a waste of his time by "bureaucrats".

The householder method is obviously less expensive because it requires a smaller number of staff, and no elaborate training programmes for interviewers are necessary. Again where the information desired is considered very personal or confidential, the use of the householder method may generate some degree of co-operation or response from those who may be unwilling to discuss such questions with "strangers" (interviewers), but do not mind recording the information themselves.

The major disadvantages of the householder method however is that it requires a certain amount of education to fill in the answers and hence cannot be used in countries or areas where the rate of illiteracy is very high. Even where the rate of literacy is very high, one cannot ensure uniformity in the recording of the information. Questionnaires may also be lost, mutilated or stained by careless or irresponsible householders, even if sactions are prescribed for such behaviour.

.../24

For obvious reasons, the questionnaires in the householder method has to be short and simple; this is an obvious disadvantage in that a great deal of money and effort may be spent to collect only a small amount of information.

From the discussion above it is obvious that in most African countries where illiteracy rates are very high, locations or houses are not properly arranged or addressed, and where representatives or directives from government are often viewed with suspicion or disinterest, the Canvasser method is much more practicable than the householder method. It is possible however to use the latter for special segments of the population or for special surveys (e.g. urban surveys), or even to combine both methods in the same survey for different segment of the population.

Another variation of the householder method is the use of the postal services or the mail to distribute the questionnaires rather than accredited agents of the census or survey organizations; this method is obviously less expensive since very few census or survey interviewers are needed, but in African countries where houses are not properly numbered, where the postal services are very rudimentary, and where the level of illiteracy is very high, the use of the mail method cannot be seriously entertained, except again for some special categories of persons or institutions.

TYPES OF QUESTIONNAIRES

Depending on the resources and objectives of the survey or census, it is possible to use three different types of questionnaires. These are: a) Single Individual Questionnaire b) Single Household Questionnaire and c) Multihousehold Questionnaire.

If only a few basic questions are being investigated, then the multi-household type of questionnaire is more convenient. Depending on the number of questions or the extent of detail required, the single household or single individual questionnaire may be used. For example, in a survey of the maternal history of women, the single individual type may be more convenient for checking, compilation and sorting. This will mean however that one has to print and handle more questionnaires than in the other two cases.

Pre-coded questionnaires are sometimes advantageous but may result in loss of information, if plans were not properly made at the early stage of preparations.

ENUMERATION OF SPECIAL GROUPS

Certain categories of people often present problems to the census organizer owing to the nature of their life-style or mode of living. Whether a de-facto or de-jure method of enumeration is used, a vast majority of the population can be contacted relatively easily at their places of usual residence for the interview. But a few other people

.../25

are not "settled" in regular dwelling units. Examples of these are (a) those in institutions such as hospitals, boarding schools, army barracks etc. (b) the floating population in institutions such as hotels, remote camps etc. (c) Transients, homeless people or outdoor sleepers (d) Nomads and semi-nomads, and (e) Special groups such as pygmies or isolated religious communities.

Normally where a de-jure method of enumeration is used, it has been assumed that many of these categories of people such as those in hospitals, schools, hotels, would be listed or enumerated in the households where they normally reside since their stay in the institution is temporary. Where a de facto method of enumeration is used however, these categories of persons present a serious problem owing to their large size, and it is often necessary therefore to use special techniques to enumerate them.

For those in large institutions such as boarding schools, prisons, camps etc., the usual practice is to enlist the assistance of the heads of these institutions to undertake a complete pre-listing of the inmates prior to the census night; one or several enumerators may be assigned this task, and then a quick check is made on census night to ensure that there have been no additions or deletions since the list was compiled.

Since the population of such institutions as hotels, lodges, motels, etc. are very unstable, the usual technique is to ensure that the enumeration of such institutions can be completed in one night. Similarly, it is also important to complete the enumeration of all such floating populations, outdoor sleepers, transients etc. in one night to avoid duplication and omissions. For those who are actually in motion on census night and could not therefore be covered by this special one-night enumeration, - such as fishermen, hunters, those travelling in the air - the general practice is to relax the strict de-facto principle and enumerate them at their previous or next points of call.

In those countries such as Libya, Mauritania, Somalia and Sudan with large nomadic populations, the enumeration of the nomads always presents special problems since the nature of their social life makes it difficult to locate them at a precise point at any point in time.

It is necessary here to distinguish between pure nomads who are constantly on the move with their families and livestock in search of pastures, and semi-nomads who operate within certain well-defined zones, and engage in agriculture during certain periods of the year. It is obvious that the former would be more problematic for the census organizer than the latter.

Apart from the problem of constant mobility such groups are generally cut off from the main stream of national life and very suspicious or hostile to outsiders especially government agents. It is therefore always important to let either their own people or officials who normally deal with them to assist or undertake the enumeration.

.../26

In general six main techniques have been used by various countries to enumerate their nomadic populations. These are: (i) Water-point approach; (ii) Camp approach; (iii) Enumeration area approach; (iv) Group assembly method; (v) Aerial Survey method; and (vi) Tribal or hierarchical. approach.

The first method involves a complete pre-listing of all the water-points in the nomadic region around which the nomads periodically congregate and then sending enumerators to all these areas to undertake the enumeration before they disperse again. The major problem in this method however is that it is not always easy to obtain an accurate list of the water-holes, and as the populations vary widely from hole to hole, the interviewers may well find that half their respondents have vanished before they have completed their enumeration.

Again quite often only children and minors who normally take care of the herd congregate at the water-holes, whilst the adults may be camping some distance away from the water holes.

The camp approach is similar to the water-hole method in that it involves a pre-listing of all the nomadic camps in the region. It also suffers from the same major drawback, that is, it is often difficult to get an accurate list of the camps.

In the enumeration-area approach, the country or district is divided into conventional enumeration areas; thus each interviewer is expected to locate in the normal way all nomadic or semi-nomadic groups within his enumeration area. This means that a large number of interviewers are needed to complete the enumeration in the shortest possible time to avoid duplications and omissions.

The group assembly method involves the use of the leaders or heads of the nomadic communities to persuade their people to assemble at pre-arranged locations for the interviewing. This involves considerable pre-census preparations and publicity in order to secure the co-operation of the tribal leaders. As the leaders have limited powers over their own particular class or lineages, it is not always easy to determine and estimate the size of those who fail to turn up for one reason or the other.

The objective of the aerial survey method is more to estimate the size of the nomadic population than to obtain any information on their social and economic characteristics. It involves the counting of the number of camps and tents from the air and multiplying the number of camps and tents by the estimated average number of persons per tent to obtain the size of the nomadic population. It may be combined with the actual enumeration of a sample of camps or tents in order to obtain some information on the social and economic characteristics of the nomads.

The last method, which seems to be fairly popular in countries with sizeable nomadic population, is the tribal or hierarchical method. The nomads normally live in organized social groups under tribal chiefs or

leaders whose authority and power vary depending on the position of that particular group in the hierarchical lineage structure.

The hierarchical approach therefore basically involves the use of the tribal leaders to collect the desired information. This may be done in two ways; one is to collect all the desired information directly from the chief, whilst the other method is to seek his assistance in collecting the information directly from his subjects.

The success of this method depends a lot on the popularity of the chief and the degree of influence or control which he exerts on them. The complete dependence on him means however that if he is not very co-operative, or forgetful or has no real control over his scattered people, then the results may be adversely affected.

For other categories of "difficult" people such as isolated religious communities, pygmies, etc., it is often necessary to see intermediaries such as missionaries, trusted chiefs or leaders to persuade them to co-operate with the census team.

2.3 AVAILABILITY OF DATA - NATIONAL PUBLICATIONS AND INTERNATIONAL AGENCIES (U.N. AND OTHER)

The work of a data collection agency does not end with the collection and compilation of the information. It extends to transmitting the information to the users in a usable form. Thus publication and dissemination of the data are important functions and hence great care should be taken to achieve these objectives so that they serve the maximum number of users. In the early stages of planning of the data collection, the focus should already be set on the types of publications which will be brought out. This may get modified after the data have been collected and processed, due to one reason or the other. In many cases, the data users would be consulted at every stage and the types and varieties of questions, the formats and depth of the tabulation and the types of evaluation and analysis which would be carried out may be planned in advance. Adaptations may become necessary due to the data and other reasons. However, efforts should be made to bring out publications covering the various aspects of data collection in a clear and user oriented manner. Thus, in addition to tables, there would be need for explanatory notes regarding how the data was collected, the problems encountered and how they were tackled, the lessons learnt and recommendations for the future. Then there should be publication on evaluation and analysis of the data in a clear and concise form understandable to the ordinary user. Technical appendices should be relegated to the end of such publications.

However, it may not be necessary to print all data. Some of the detailed information may be made available in limited numbers to special users. Computer print outs, special tapes or discs etc. have been used for such purposes.

.../28

Also the detail of tables may vary from the national to the regional, local and other minor subdivisions. Sometimes publication is by subject matter speciality, viz demographic characteristics or economic variables etc and sometimes it would be by geographic coverage. And it is quite obvious that the greater the number of geographical areas for which the data are needed, the larger the size of tables and materials and since accuracy and reliability may become less and less as the area units become small, it may be advisable to produce such detailed information only in condensed fashion.

The question as to whether tables should present only the raw data or also some derived statistics like mean percentage etc the main recommendation is that it is advisable to present raw data. However, if some preliminary analysis would aid its utilization and interpretation and would not mutilate it, then it is advantageous to present these also. Sometimes, for saving time and space it may be necessary to present only derived parameters viz, percentages but then it is essential that the total numbers on which the calculations are based may be necessary. In some cases, unless warranted, condensed tabulations may suffice. For example, in the table on children surviving it may be sufficient if the total number (by sex) is tabulated by age of mother instead of giving mothers by age and number of children surviving (0,1,2...). In the latter case, there should be a column for total number of children surviving unless the last open interval does not contain a high proportion of cases and the calculation of the total can be done by computers. Some of the publications from Tanzania are indicated in section 1.1.

In addition to national agencies, several international agencies are interested in comparative statistics. Chief among these are the U.N. agencies like the U.N. Statistical, Population and Social Divisions, the Food and Agricultural Organization, World Health Organization, the United Nations Educational, Scientific and Cultural Organization, the International Labour Organization etc. These organizations collect and compile demographic and socio-economic information on a global basis and publish the data in annual or regular series. The Statistical, Demographic, Production and Trade Year Books, the Year Books of Labour Statistics, and Educational Statistics are yearly publications dealing with specific topics like population, labour force, education, food production, trade etc. Then there are other regular publications like the Compendium of social statistics, Public Health and Vital Statistics Reports etc. which pertain to living conditions (housing, transportation, communication etc) and health and vital statistics.

Some other international agencies like the U.S. International Statistical Program, the Organization for Economic Co-operation and Development etc. also compile data on an international basis and these are valuable sources of comparative data. They also many times, provide some analysis of the information.

However a word of caution need be sounded in respect of data from international sources pertaining to timeliness, coverage and reliability.

These being mostly secondary sources, it is advisable to rely more on primary sources like national publications especially if one is looking for recent data in some detail. But for comparative studies, such international sources are invaluable.

3. DATA PREPARATION

3.1. FIELD CHECKS AND SUPERVISION, EDITING, SCRUTINY AND CODING

A census or survey is usually a very large and complex operation involving a very large number of various categories of personnel. There are firstly the technical officials directly responsible for the collection of the statistical data. Though the designations used for this category of officers vary occasionally according to the administrative structures of different countries, these officials may range from the enumerator in the field who actually collects the data, through his team leader, supervisor, district supervisor, regional census officer and finally to the chief census officer who is normally assisted or advised by a census technical committee.

Apart from these technical officers, no census machinery is ever complete without a large number of supporting staff such as drivers, administrative and accounting staff, technicians, data processing specialists, public relations practitioners etc.

For such a large number of people with diverse backgrounds, training and expertise to function effectively for the success of the enumeration, it is essential for a clearly defined organizational structure to be set up with clearly demarcated lines of authority and responsibility.

One implication of this hierarchical structure is that each technical officer is directly responsible for the work of the officers below him. Thus while the work of the enumerator is constantly checked by his supervisor in the field, the supervisors in a district also in turn constantly report to his superior officer. Field checks are especially important in the early stages of a census operation, since it permits necessary changes to be made early in order to ensure that errors fall within acceptable limits.

For example, in spite of all the training and pre-testing, one may find that responses to a particular question are unsatisfactory, and new instructions may then have to be issued immediately to all interviewers. Some interviewers may be found to be irresponsible, drunk, sick or absent, necessitating urgent remedial measures. Thus one category of officers has to constantly monitor the work of those under his authority in the field to ensure the success of the effort. Such monitoring relates not only to the quality of the data being collected, but also the safety of the documentation (mainly questionnaires).

In spite of these checks, the short period within which a census or survey has to be carried out does not always make it easy for major

.../30

errors to be corrected. It is therefore necessary after the census has been completed or when it is in final stages, for the supervisor or district officer to make a quick check or scrutiny of the questionnaires and other documents he has received both with respect to the quality and quantity of the data. If the quality falls below certain pre-determined limits or if some documents are missing, the work may then be sent back to the interviewer concerned for immediate corrections to be made.

The scrutiny may be done at various stages of the operation, but when it is done in the central office after all the questionnaires have been collected, it is usually then too late, or too expensive and cumbersome to return the questionnaires to the field again. It is therefore more advisable to do it at the local level just at the end of the operation.

EDITING AND CODING. In order to process the data collected, it is necessary to transform the respondents' recorded answers into numeric symbols which can be read by the machines (usually computers) which have been specially designed to process the data. The use of such numeric symbols to represent the data collected is called coding, and represents one of the most important stages in the census process.

In most censuses, it is usual to pre-code certain responses which are relatively easy to obtain such as sex and marital status. Such pre-coding is not common or advisable however, where the range of possible alternative responses is large e.g. occupation. But even here, one may decide to pre-code occupation into about ten or so categories, if one is not interested in a sophisticated analysis of occupational structure.

The main advantage of pre-coding is that it saves time; it however suffers from one major disadvantage. When a large number of items or questionnaires have to be pre-coded, interviewers sometimes become bored and mark the wrong boxes without even being aware of this. Forcing them to write down the whole answer tends to eliminate such possibilities.

There are two main coding techniques. One is to allow a coder to code all the items relating to a particular respondent; the other is to allow a coder to 'specialize' and code only part of the items e.g. demographic characteristics, economic characteristics, maternal and fertility history etc.

The specialization technique permits easier or quicker familiarisation with the codes and is therefore faster. The comprehensive approach is more laborious and time-consuming. However it makes it possible for the coder to detect inconsistencies in the responses. For example, if a respondent aged 12 is recorded in the occupational column as being an engineer, the coder knows there is a mistake somewhere which has to be corrected. It is not possible to detect such an error in the specialized approach since the two items will be coded by different people.

.../31

An important part of the coding operation is verification of coding. This may be done either on a sample basis or on a hundred per cent basis. It is more common however to start on a hundred per cent basis, and then change gradually to a sample basis after the coders have acquired various levels of efficiency.

Since the scrutinization of questionnaires from the field is usually very quickly done, it does not lead to the elimination of all errors in the questionnaires. Some responses may be inconsistent with each other, while some columns which should be filled may be empty. The questionnaires therefore have to be corrected before they are processed. Editing therefore simply means correction of the data in accordance with predetermined rules. The editing may be done at various stages of the operation. It may be done by the interviewer or supervisor in the field or in the office before or during the coding process.

It may be done either manually or by machine, and various software packages (such as CONCOR, UNEDIT) have been developed to facilitate speedy editing of data.

3.2 CODING, EDITING, PUNCHING, VERIFICATION AND DATA TRANSFER

Editing and coding are key steps in any survey operations. They constitute the link between the raw data collected from the field by the interviewers and the input to the computer programmes.

Editing is a process designed:

(a) to check that information contained in the questionnaire is complete, recorded in the prescribed manner and consistent; and

(b) to take appropriate action when these conditions are not fulfilled.

ORGANIZATION CHART

Ideally, an organization that can handle the survey data already exists in the country and only the proper channels of command need to be established. However, additional training and/or personnel may be needed to supplement the existing organization.

DOCUMENT FLOW

Figure 1. shows steps that are required from the time questionnaires reach the office from the field until they are handed over to punching unit.

SYSTEM OF ORGANIZATION

(a) Fixed system - where one person has been trained for or assigned to, a specific job throughout the operation and does nothing else.

.../32

(b) Floating System - where a person has been trained to do a variety of jobs and alternates between, say, editing transcription and checking or between coding and verification.

PERSONNEL

The organization will consist of:

(i) Office processing Manager; (ii) Supervisor - (a) Editing (b) Coding; (iii) Receipt and control clerks; (iv) Editors; (v) Editor - Verifiers: (vi) Coders and Transcribers; and (vii) Verifiers of Coding and Transcription.

FLOW CHART FOR DOCUMENTS

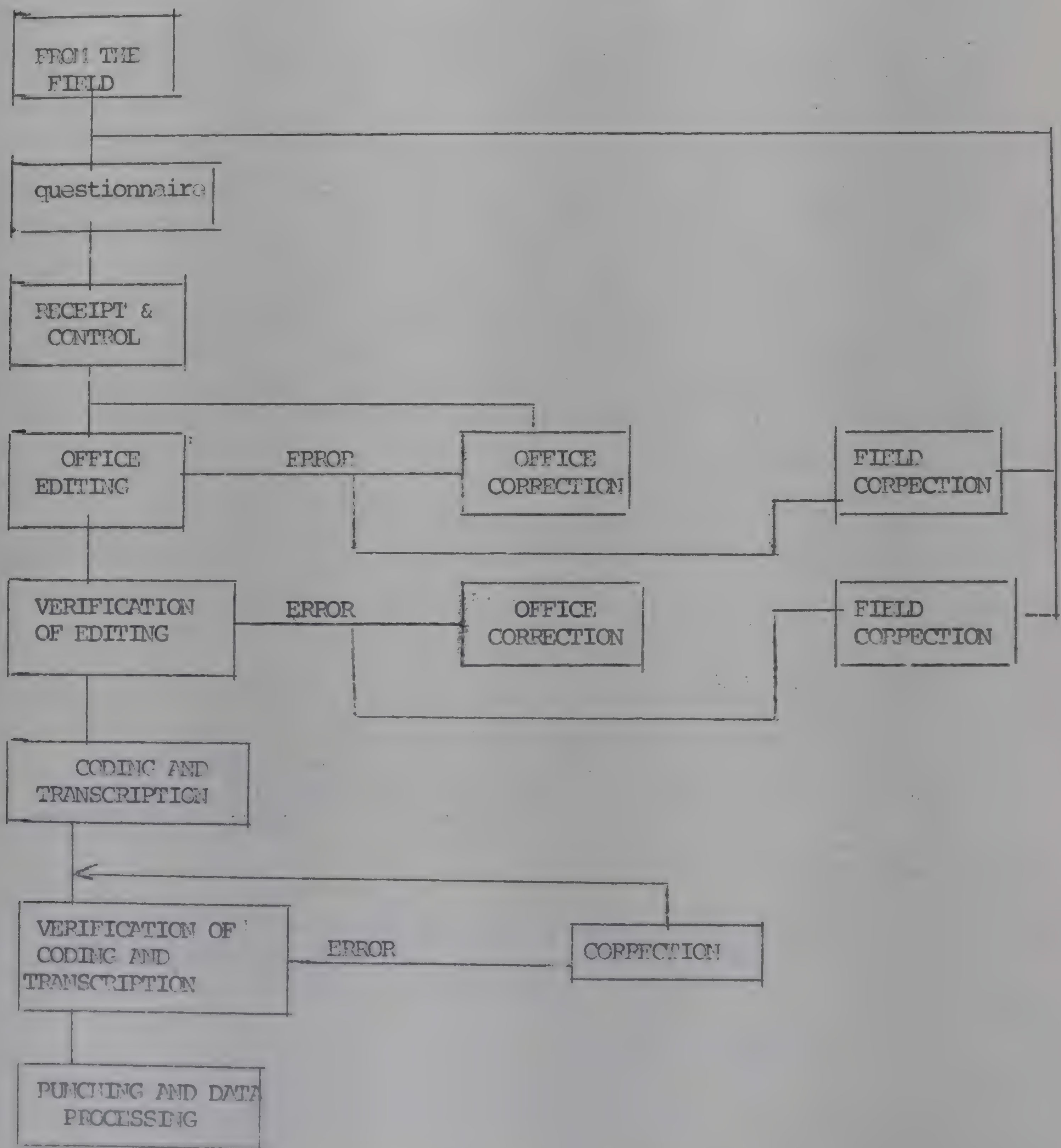


FIGURE 1.

PUNCHING OF DATA

After the data has been coded and edited properly further preparation of data before actual processing is undertaken, is necessary. This stage must be organized to ensure that all transactions entering a system are correctly transcribed into some medium readable by computer. The common media are punch cards, paper tape, Discs and magnetic tapes. Transcription of data into cards is achieved by an operator using a key-punch. The operator reads the original document and enters the characters read on to a keyboard which automatically punch appropriate holes into the input medium (cards). The process has several potential source of errors such as the illegibility of characters on the original document or the operator may depress a wrong key or may read a character incorrectly.

VERIFICATION OF DATA

In order to reduce number of errors as a result of key-punching a form of check known as verification is essential. In case of punched cards a further operator takes original documents and the cards punched from them and feeds the cards again - through a verifier. The cards to be verified and the source documents are all kept in the same sequence and the verifier operator repeats the keystrokes made by the original operator. However on this occasion no holes are punched into the cards but a sensing operation takes place to ascertain the holes previously punched agree with the keystrokes. Any disagreement will be shown by the warning lamp of the keyboard. This error must be corrected before the operator proceeds on to the next card.

DATA TRANSFER

The transferring of data from original source documents on to media readable by computer is known as Data Transfer. Some of the media used have already been mentioned above.

DATA PROCESSING

Before processing of data is undertaken, a thorough check of each record is carried out by the Computer. A validation programme must be developed in order to catch those records that do not meet rules specified for a particular Survey/Census. The rules will cover such things as range checks for various items and consistency checks within the record as well as within, say, the household.

Example of range check could be days of the month which should be between 01 - 31 or months between 01 - 12. Example of consistency check may be: Age of mother Versus age of her children; Education Versus age etc. Through this programme a master file is created for all those records which pass prescribed tests while those records which fail must be rejected so that they can be corrected and re-input into the programme to ascertain completeness.

.../34

BACKING STORE

Main store (central processor) is very fast and efficient, but also very expensive. To increase storage capacity, the main store can be supplemented with slower but cheaper devices. This additional storage for data and programmes not currently required is called Backing Store. The devices used to provide backing storage are magnetic storage devices which work on the same principle as tapes used for sound recording. The following is a list of some of the backing storage devices:

(a) - Exchangeable Disc Storage (EDS); (b) - Fixed Disc Storage (FDS); (c) - Fixed and Exchangeable Disc Storage (FEDS); (d) - Magnetic Drums; (e) - Floppy disc (diskette); (f) - Magnetic Tapes; and (g) - Cassette Tape.

3.3. DATA PROCESSING AND RETRIEVAL

FILE ORGANIZATION

There are three methods of setting up a computer file:

(i) Serial organization; (ii) Indexed Sequential organization; and (iii) Random organization.

FILE ACCESSING

The order in which records are retrieved from a file held on a direct access device is called the file access method. There are four methods of file accessing:

(i) Serial Access

- Every record is ready in turn as it is encountered, starting from the beginning of each seek Area and processing Seek Areas in turn from the beginning of the file.

(ii) Sequential Access

- This method makes sense when applied on a file stored sequentially. The housekeeping overflow routines ensure that records stored in the overflow buckets are presented in correct key sequence. Sequential access can be "selective" or "non-selective".

(iii) Selective Sequential Access

- This method improves efficiency if the hit rate is so low that the number of records being skipped over is more than the total number of records in the entire file.

(iv) Random Access

- Means extracting a particular record without searching the whole file serially. A method known as address generation technique is used to achieve this. Where the file activity (hit rate) is very low, this method should seriously be considered.

PROCESSING MODES

There are three processing modes on direct access devices:

(i) Input mode; (ii) Output mode; and (iii) Overlay mode

(i) Input mode

In this mode records can be read from Disc to store but no transfer from store back to disc is possible. See Fig. 2.

(ii) Output mode

In this mode, records are written from store to Disc. It is normally used for creating serial files. See Fig. 3.

(iii) Overlay mode

In this mode (not possible on Magnetic tapes) the bucket is ready in store, updated in store, and then written back to the same physical position on the catridge it came from. This is the processing mode used when updating in situ and sometimes called Input/Output mode. See Fig. 4.

ADVANTAGES OF USING DIRECT ACCESS DEVICES OVER MAGNETIC TAPES

1. In real time system - response time is critical. Apart from speed, Direct Access allows selective file processing.

2. When the proportion of data being accessed on a file is low, it is more efficient to go directly to the required item than working serially through the whole job.

3. Avoid time-consuming complex SORTING.

Fig. 2 REPORTING

Fig. 3 UPDATE BY COPY

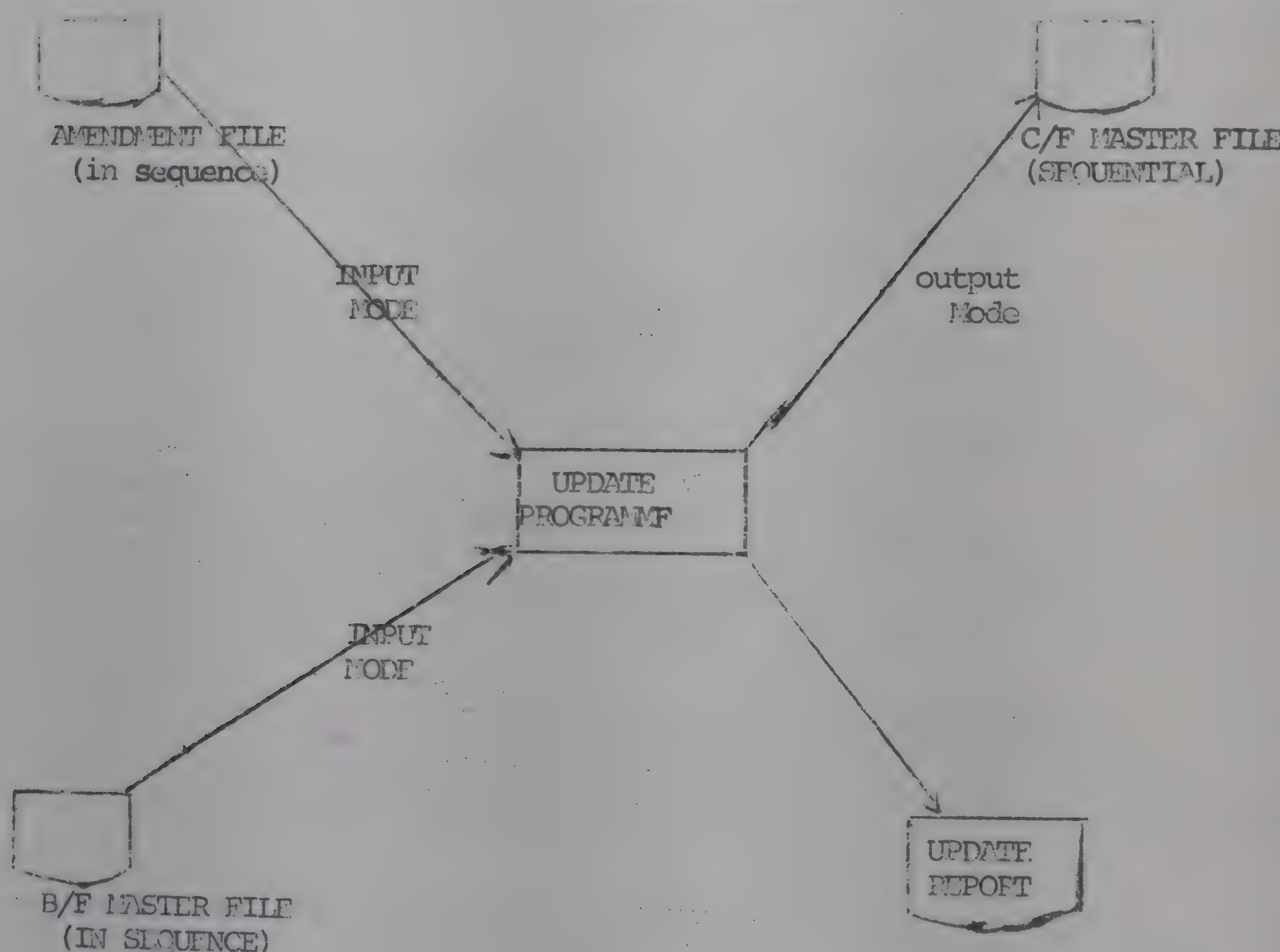
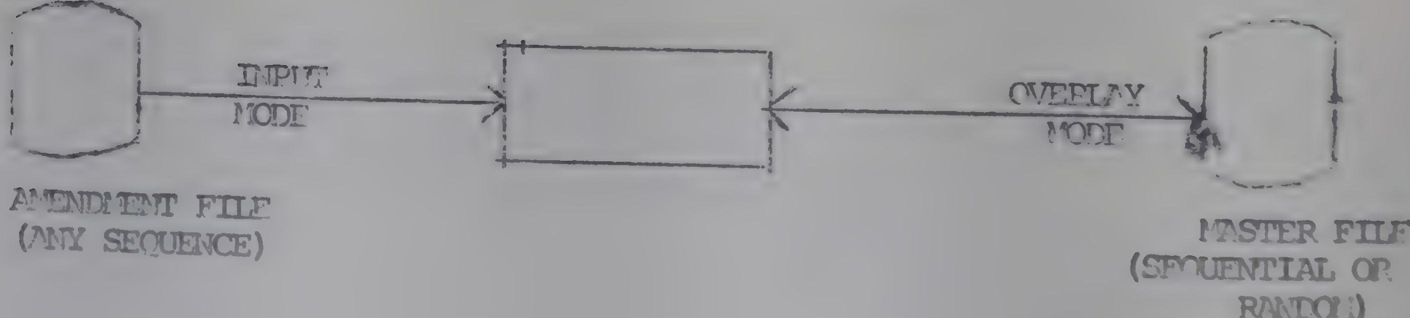


Fig. 4 UPDATING IN SITU



BASIC REQUIREMENTS OF COMPUTER INSTALLATION

1. Personnel
2. Computer (Central Processor) and its peripherals
3. Programmes
4. Data

1. PERSONNEL

- Administrators/Engineers; Systems Analysts and Designers; Programmers; Operators; Data Control Staff; and Supporting Staff.

2. CENTRAL PROCESSOR

The processor has three separate hardware sections:

(i) Internal memory:

- used to hold data or programme instructions

(ii) Arithmetic Unit

- consists of special registers and circuits which are able to perform arithmetic and logical operations upon operatiors selected from memory. These operations include addition, subtraction, division, multiplication, comparison and logical "OR" and "AND" capabilities.

(iii) Control Section

- This section has complex functions to perform. The section co-ordinates the activities of other sections of the central processor including the operation of peripheral units.

3. PROGRAMME

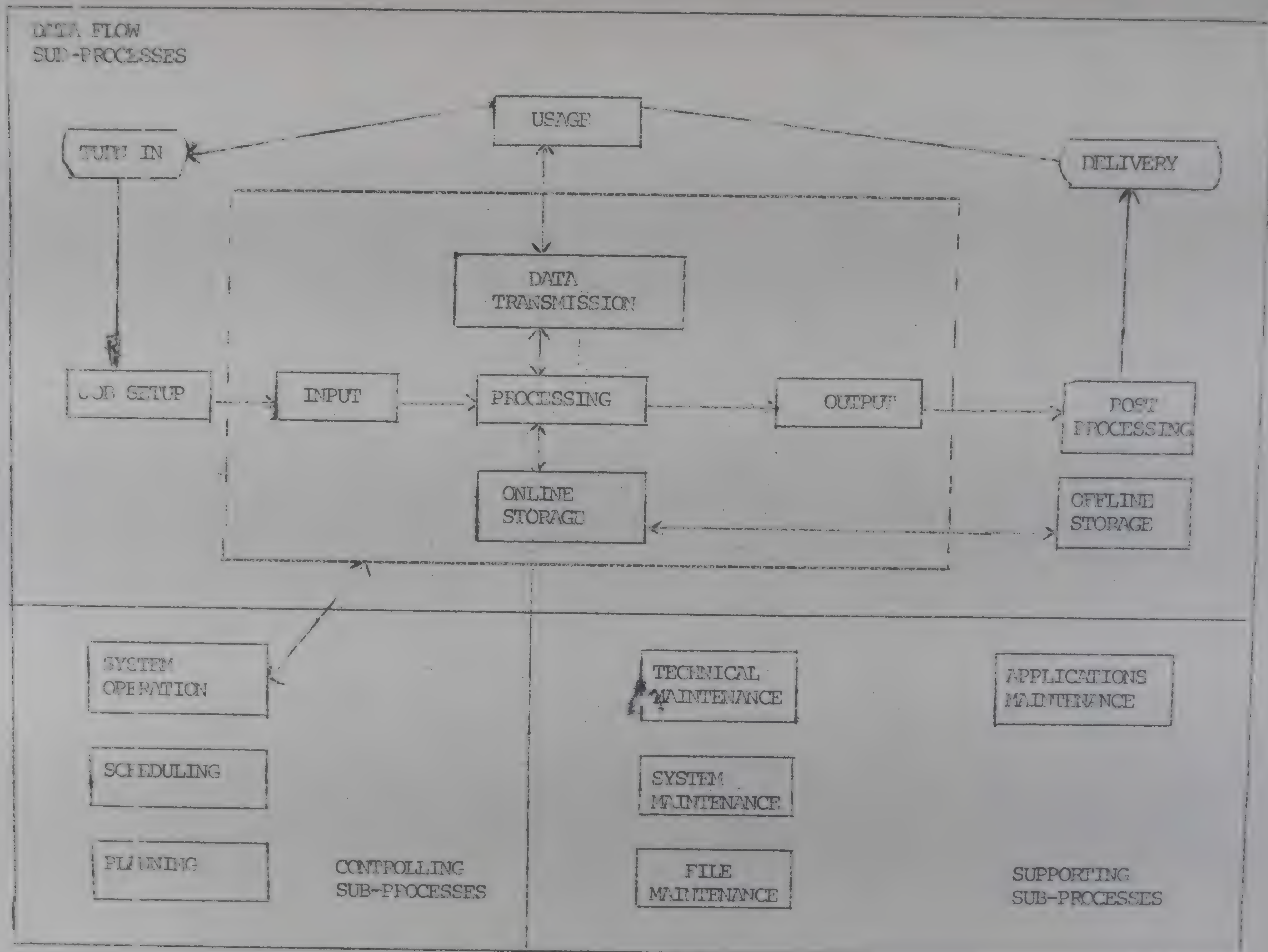
- A programme is a network series of instructions telling the computer how to carry out a particular job.

4. DATA

- Means straight forward facts and figures which the computer needs to work on.

Figure 5. Below shows relationship between various sub-processes in a computer installation.

Fig. 5


DATA FLOW
SUB-PROCESSES
USAGE
TURN IN
DELIVERY
DATA
TRANSMISSION
JOB SETUP
INPUT
PROCESSING
OUTPUT
POST
PROCESSING
ONLINE
STORAGE
OFFLINE
STORAGE
SYSTEM
OPERATION
SCHEDULING
PLANNING
CONTROLLING
SUB-PROCESSES
TECHNICAL
MAINTENANCE
SYSTEM
MAINTENANCE
FILE
MAINTENANCE
APPLICATIONS
MAINTENANCE
SUPPORTING
SUB-PROCESSES

3.4 TABULATION - FORMAT, TYPES AND DETAILS

The mass of data produced in any data collection exercise becomes meaningful only when they are condensed into summary statements or tables. As a matter of fact, tabulation of the data is the first step in the evaluation, analysis and interpretation of the data.

In any enquiry, once the focus is set on the types of data needed, it is necessary to prepare table outlines and dummy tables. This ensures that all the data needed for the tables are collected and that unnecessary data are not collected.

Ideally, the questionnaire and enumeration manuals are finalized on the basis of the analytical and user needs which in turn determine the types, depth and varieties of tables needed. Data producers and users will have to decide on these at the earliest stage of the exercise, so that, there will be no bottleneck in the end.

Tables may be classified into various priorities according to the importance they will have in their utilization. This also determines the timing and sequence of their preparation and the details into which they will be presented. Space in the publication, time, finance, usefulness of the information and quality of data etc. will determine the breadth and depth of the tables. For example, national tables may go into great details on various demographic, socio-economic characteristics but at the local or small areas level, most of the information may be presented only in summary form or for grouped values. This may be necessitated by the small samples involved also.

Presentation of tables could be by geographic areas, subject specialities or otherwise. Decision on a particular format will depend on circumstances.

It is advantageous to present the data as it was collected without any alterations, adjustment or analysis. However, if any evaluation or analysis is to be presented, then the user must have recourse to raw data and an indication of the techniques and methods applied, so that independent conclusions can be drawn, if necessary. In some cases, for saving time and space, it may be expeditious to present only percentages, means etc. In such cases, it is important that the numbers on which they are based be provided. On the other hand, in specific instances, analysis may not warrant presentation of data in the details in which they were collected. For example, data on children ever born is collected under three headings - children living here, those living elsewhere and those now dead. The information in such details is useful for data evaluation and adjustment especially in the field and office. For presentation of results, it is sufficient if tables are provided by age of women and total numbers of children ever born (by sex, if possible). If parity distribution is to be presented, there should be a column giving the total numbers of children. Otherwise, it is tedious to calculate the weighted sum and the problem of open classes also will arise. This is more so in the presentation of data on children surviving.

.../40

In other words, the aim of tabular presentation should be (1) to be precise and clear (2) present all information needed but not anything else and (3) digest the data for evaluation, analysis and interpretation.

Not all tables may be printed or published. Some of the tables may be only for limited use and made available in computer print outs etc. In some cases, special tapes may be prepared (with identification and other details removed) for research purposes. It has also become acceptable in many countries to provide sample data tapes for study outside the country. Special tabulations also are provided, on request by genuine research bodies or individuals - sometimes free but other times at nominal costs.

4. DATA QUALITY

4.1 IMPORTANCE OF QUALITY CHECKS AND CONTROLS

Introduction

Whether one collects statistical data using the complete census method or the sampling method, no one has ever been able to have complete information. There has always been a difference between the information collected and the true information. But, since no two data collection exercises can ever be replicated, none has ever known what the true information is for any population. Also, no two data collection exercises on the same population have produced exactly identical results. We have called the difference between the true information and the information collected as an error or bias. When the information is collected by sampling method, another error is introduced because not all units of the population are observed. Thus we have what we call sampling errors and non-sampling errors.

While sampling errors are taken care of by choosing the right sampling schemes and the right sampling sizes and using the relevant statistical formulae, non-sampling errors are usually taken care of by quality controls and checks. Control checks would test whether an interviewer in fact made all the interviews claimed, whether his response rate is satisfactory and whether he is asking the questions and interpreting the answers in accordance with instructions.

Non Sampling Errors

(a) Coverage errors

The first class of non-sampling error is where there is an error in coverage. This may be in the form of omitting a unit which is supposed to be included, including a unit more than once or including a unit which should not have been included. To avoid these errors, control checks are introduced right at source where the data are collected. The immediate supervisor of the enumerator is required to check that all forms returned have been

filled in completely and by the right respondents. Sample checks for each enumerator may help to determine whether the right respondents have been covered. Comparison of results and lists from other sources is another way of checking. Once the forms have been returned to the head office, to determine the degree of error, post-enumeration sampling checks are carried out.

(b) Non-response

This arises from incomplete coverage of the units qualifying to be included because either the units cannot be contacted or the units are contacted but no observation has been made because of refusal. To correct this type of error, the ingenuity of the immediate supervisor is called for in either locating the unit or convincing the respondent to respond. If the percentage of non-response is very high, then the results of the analysis on the data collected is questionable.

(c) Response and Processing Errors

This consits of errors in recorded value and include, in addition to response errors those of editing, coding, punching and calculations. Response errors are sometimes more important source of bias than non-response. It arises from a variety of causes including the faulty memory of the respondent and the conflicting reports that may be secured from different members of a household. Various methods are in use for correcting such biases, depending on the nature of the study. The availability of external data for comparison and the accuracy required is one method. The method of post-enumeration surveys is also useful.

(d) Estimating Error

This is mostly due to the use of wrong or biased formulae for estimation. The amount of the bias introduced by such methods are usually estimable and should not present a serious problem. If the bias is found to be large, one can avoid using the method.

(e) Enumerator bias

In the case of personal interview method, the fact that different enumerators are used for different enumeration areas, results in the lack of uniformity of treatment. Ideally, respondents are supposed to be treated uniformly when being asked questions. Differences in respondent treatment may be a source of error. This problem is partly tackled by the enumerator training and partly by the checks by immediate supervisor. Also uniform application of methods of interviewing and recording is treated in a similar manner.

.../42

Conclusion

Whatever the preparation that may be made in the office before the data collection exercise commences, there is always room for errors to creep in. The only way of reducing these errors is by applying the quality checks and controls.

4.2 METHODS OF QUALITY CONTROL

The importance of assuring the quality of data needs no stressing. Data collection is expensive and can be done only infrequently. Wrong data can lead to wrong decisions which in turn can result in waste of resources, time and upset plans and policies. The confidence of the people in general and the users of data in particular will be lost if the data made available are not of good quality. Since data are used for planning and policy formulation involving estimation and projection, base data errors could get compounded and confounded, resulting in biases and lop-sided decisions and action programmes.

Unlike that in physical and exact sciences, it is not possible in the social sciences to obtain precise information nor is it possible to repeat the exercise under identical conditions. Hence, a certain amount of error or bias is inevitable. What one tries to ensure is that their magnitudes are not too large and that they are minimized. Again, in an enterprise like data collection involving several steps and stages and human and mechanical agencies, it is quite possible that errors, biases, defects and deficiencies could creep in. Quality checks and controls are devices to gauge these in the data and take appropriate measures to contain them.

Statistical data are subjected to sampling and non-sampling errors. Good sample design, large sample size and care in data collection can reduce sampling errors. But increasing the sample size without improving the data collection machinery and taking other precautionary steps can result in increased non-sampling errors. A well designed and executed small sample survey is superior to a badly designed and conducted large scale enquiry.

Non-sampling errors could be (1) Coverage errors and (2) Content errors, i.e. error arising from problems in the coverage of the study population and errors in data covered in the enquiry in respect of characteristics etc. Coverage errors could be independent of content errors but may bias the results, if coverage is selective or otherwise non representative. At the same time, coverage error may occur because of content of questionnaires: similarly content error could be selective of areas or group and may result in coverage error. These will be considered in later sections.

Prevention is better than cure and hence it is a wise policy to anticipate the possible avenues for errors, biases etc. and to plan for their minimization, if not their elimination. The goal of checking and controlling the quality of data is to improve quality before

data are collected and minimize loss of quality after it is collected.

To ensure a real coverage, geographic preparation should be adequate and demarcation of enumeration areas clear and unambiguous. Enumerators and supervisors must know their areas and advance publicity of their activities will ensure better co-operation from the population. Special problems like remote and inaccessible areas, one person households, migrant group, nomadic or semi-nomadic population, infants and young children, females or males in specific ages etc should be specifically catered for.

Also great care must be taken to ensure that questions pertaining to specific groups like those aged 12 years and above for marital status, fertility questions to females aged say, 10 or 15 years and above, economic activity for those aged above 12 years etc. do not result in coverage problem. Strict rules for inclusion/exclusion of members of households in enumerated households should also be spelt out and should be such that on the whole there is no omission or duplication.

If there be different types of questionnaires for various segments of the population, it should be ensured that these are adequate for the work and canvassed properly as per plan. This is specially true of long detailed questionnaires usually meant only for a sample of the population.

The questionnaire containing a large number of items covering diverse subjects should be pretested in the field after having been thoroughly discussed by users and subject matter specialists. Especially this is important in cases where the language of the questionnaire is not the same as the local language/languages. Again, difficult concepts like economic activity, fertility, etc. should be clearly spelt out. This leads us to the need for training of enumerators and supervisors not only in the class room but also in the field (by mock interviews etc.) and educating the public. Good supervision and continuous monitoring of the data collection will ensure better quality of data. In the field, it is advantageous if supervisors verify the work of enumerators on a 100% basis for the first few days of the operation and discuss with individual enumerators the types of mistakes etc. they are making. Common problems may be discussed in group meetings. Problem in the field should be solved before they become massive. Enumerators should be encouraged to discuss with supervisors their problems and suggestions for their solution. It should be ensured that enumerators cover their areas and that information is collected on all members of households who are eligible for inclusion but exclude those ineligible. No question in the questionnaire should be left blank - if a question is inapplicable to an individual this should be indicated by an appropriate symbol (N.A). Even in cases where the respondent may not know the exact answer, every effort must be made to estimate or arrive at it because it is easier to do this in the field and virtually impossible after the data has been brought to the office. Inputation and other methods of assigning answers to questions become questionable and should be avoided as much as possible. Strict control of the filled in questionnaires is required, if one wishes to avoid loss or mutilation of the collected data.

.../44

In the office processing of data, i.e. editing and coding, great care should be taken to ensure that the ultimate objective of getting the types of tables etc. are kept in focus. It is advisable to verify the work of editors and coders to vouchsafe quality of work. This is necessary also for the further stages of data processing like punching and data transfer to card, tape or disc.

There are several ways of checking accuracy of data collected. Spot checking is one such but it is not recommended unless time and cost considerations rule out the use of more scientific and systematic techniques. Sample verification could be either continuous or acceptance sampling. The former is useful when the units are reasonably homogeneous in nature and homogeneously produced. Acceptance sampling is applied to tasks that are people oriented such as field review of enumeration's work, manual editing and coding and data keying. Many times complete verification may be needed at least until confidence is established on data quality.

Verification can be either on 'dependent' basis or on 'independent' basis. In the former, the verifier examines the work of a producer and decides whether it is correct or incorrect. In the latter, independent verifiers repeat the process and match results. In most cases, the second alternative is superior but it requires more care in controlling the flow of work and prevent collusion.

With the availability of the computer, it has become possible to perform consistency checks and adjust data. Software packages like concor, unedit etc. are available.

Finally, a word is in order regarding the important role of management in assuring data quality. A well managed data collection exercise taking account of the various possibilities for error and biases will definitely pay dividends in terms of better data.

4.3 EVALUATION OF COVERAGE - DIRECT AND INDIRECT METHODS

Introduction

It has been observed earlier that the main objective of any census or survey is to enumerate and record the characteristics of all persons or units within the territory at the time of the survey. In practice, it is almost impossible to achieve this perfection for one reason or the other except for very small surveys. Even where the count is acceptable, some of the information relating to the characteristics may be wrongly given or recorded.

The census or survey therefore is often beset with two major sources of errors - errors of coverage where some units belonging to the universe being investigated e.g. localities, compounds, households or persons may be completely omitted from the exercise, or are duplicated and errors of content or classification where all the units are covered, but some mistakes are made in recording or reporting

of some characteristics of the units or persons in the population such as one's age, sex, marital status.

Coverage Errors

The coverage error can either be in the form of over-enumeration or under-enumeration of units. These errors may occur as a result of a combination of several factors. The major reasons for coverage errors in African censuses are:

(i) Poor or inadequate planning leading to faulty demarcation of enumeration areas. Where the pre-census mapping operations are poorly done, enumerators may find that they have directions or descriptions which bear no relation at all to the reality. In such a situation, it is not unusual for whole localities to be missed whilst some are covered twice by different enumerators; (ii) Even where the mapping has been properly done, some enumerators may misunderstand or fail to adhere strictly to instructions and sometimes lose completed questionnaires; (iii) Some respondents may deliberately mislead the interviewers either because they want to inflate the population of their areas, or because they do not want to be counted, or out of sheer ignorance; and (iv) Owing to poor transportation and communication problems, it is not uncommon in some African countries, for the existence of some communities or localities to be completely unknown to the administrative authorities. This problem is accentuated by the existence of a wide variety of cultural practices which determine the location of certain groups of people at certain times of the year. For example, the Nuer people of the Sudan completely change their residences and mode of life during the wet and dry seasons. In Ghana, some communities shift or relocate their villages occasionally either because of disputes with neighbours, in search of water or more land etc. In many African communities, some sections of the populations e.g. specific age-grades or nubile girls may go into seclusion deep in the forest for the practice of certain rites during certain periods in the year, and of course, if such periods coincide with a census or survey, the possibility of under-enumeration is very high. Again, however hard the enumerator tries, there are certain categories of people who are very difficult to locate or interview, examples of these are illegal immigrants, those who live alone or have odd working hours, the floating population etc.

EVALUATION OF COVERAGE ERRORS

It is always necessary to evaluate the results of the census in order to determine its accuracy. This is necessary not only to guide users of the data, but also to ensure that some of the deficiencies in the census operation are eliminated in any future operations. The evaluation may also provide the statistical basis for adjusting the data in order to conform more to its theoretical maximum both for planning and other statistical purposes.

.../46

For analytical purposes, methods of evaluation of census data may be classified into two main categories namely, direct and indirect methods.

DIRECT METHOD

The direct methods involve the comparison or matching of the information collected in the census with the information collected from another source such as a post-enumeration sample survey (PES) or from re-enumeration. If the PES or re-enumeration is undertaken soon after the census, it is generally assumed that the two populations in the selected enumerations or districts would be the same; and since the quality of the PES data is assumed to be better than the original census - better supervision, highly qualified enumerators etc. - the PES can be used as a basis for estimating the extent of under-enumeration or over-enumeration in the main census by matching of persons.

In practice, this type of matching of the two sets of data has been found to be extremely difficult, if not impractical, in African countries where it has been tried (e.g. Ghana 1960, Liberia 1974).

The main reasons for this failure are i) the inability to identify places, and houses clearly because of poor addressing and demarcation of units; ii) variability of names used in different contexts by many Africans; iii) inadequate preparations for the PES owing to over-concentration on the main Census; and iv) loss of support, enthusiasm or fatigue, both on the part of officials and respondents.

A variation of the PES approach is the dual record system approach which involves the mounting of an independent survey to evaluate the census, and then attempting a systematic matching of records from the two independent data-collection systems.

INDIRECT METHODS

A wide variety of statistical techniques are also available for making an indirect test or evaluation of the accuracy of the census. The underlying principle in all these techniques is however very simple, and involves testing the consistency of the census data with other existing data or within itself. The data may be either those collected for demographic purposes such as other censuses, surveys, migration statistics, vital registration data, population register records or those which exist for other purposes such as educational statistics (for evaluating particular age groups), social insurance or employment records, voters registration list, tax records etc. These are external consistency checks. In the internal consistency method, various patterns like age ratios, sex ratios, survival ratios, fertility rate, mortality rate, growth rate etc. are utilized.

Two simple indirect techniques which are often used are the balancing equation and the growth rate method.

The balancing equation technique works on the assumption that, $P_2 = P_1 + B - D + I - E$, where

P_2 is the population enumerated at second census, P_1 is the population enumerated at first census, B and D are the total number of births and deaths respectively occuring during the intercensus interval, I and E are respectively the total number of persons who immigrated into and who emigrated out of the country during the intercensus interval.

Obviously, the balancing equation method can hardly be used in most African countries where accurate statistical data on births, deaths, in and out migration are non-existent.

More feasible for many African countries with two or more reasonably acceptable censuses is the use of a modified version of the balancing equation:
$P_2 = P_1 + B - D$, i.e. assuming that net migration is negligible or nil. Since B-D is the natural growth, one can use the intercensus growth rate as an index of coverage of two censuses.

For example, Table 2 gives the growth rates by sex for the intercensus periods 1948 to 1978. The sex ratios of the population also are given.

Table 2. Total population, sex ratios and annual average growth rates for Tanzania, Mainland and Zanzibar, 1948, 1957, 1967 and 1978

Country/Area	Population	Sex ratio	Annual growth rates		
			Both Sexes	Male	Female
TANZANIA					
1948	7744600	94			
1957	9084100	94	1.77	1.77	1.77
1967	12313054	95.2	3.04	3.11	2.98
1978	17512610	96.2	3.20	3.25	3.16
MAINLAND					
1948	7380400	93			
1957	8788500	93	1.79	1.79	1.79
1967	1195654	95.0	3.08	3.19	2.98
1978	17036499	96.1	3.22	3.27	3.17

.../48

Table 2 (ctd.)

Country/Area	Population	Sex ratio	Annual growth rates		
			Both Sexes	Male	Female
ZANZIBAR					
1948	264200	110			
1957	295600	111	1.25	1.30	1.19
1967	354400	102.4	1.81	1.43	2.24
1978	476111	98.8	2.68	2.51	2.84

Source: For 1948 and 1957 statistics - Bureau of Statistics, Recorded population changes 1948-1967, Dar es Salaam, 1967.

For 1967 - Bureau of Statistics, 1967 Population census, Volume I, Statistics for enumeration areas, Dar es Salaam, Central Bureau of Statistics, 1969.

For 1978 - 1978 Population census, Volume II, Population by age and sex for villages/wards and urban areas, Dar es Salaam, Statistics, 1981.

The growth rates fluctuated from less than 2 during 1948 to 1957 to more than 3 during 1957 to 1967 and 1967 to 1978. The low growth between 1948 to 1957 is attributed to some under-enumeration in 1957. The increase in the sex ratio also is indicative of better coverage in recent years, as normally we would have expected a fall in sex ratio. It is suggested by the analysts of the 1967 census that in 1957, there may have been more under-enumeration of females than males. Since males were subject to a taxation system, there could have been an under reporting of males also. Since changes in growth rates and sex ratios should be within reasonable limits and the values themselves should lie in certain ranges, this method is a good diagnostic tool in the evaluation of coverage of censuses.

Another variation of the balancing equation wherein we consider only those already alive at an earlier census and compare them with the survivors, is based on the version of the balancing equation: $P_2 = P_1 - D$, where P_2 are the survivors of the population P_1 at an earlier count and D is the number of deaths among the population P_1. Survival ratios of cohorts are calculated and compared for evidences of omission, duplication or age-sex reporting errors. To remove some of the inherent errors due to age-sex reporting, we can use overall survival ratios. Table 3 gives both the usual cohort survival ratios and overall survival ratios by sex for the period 1968-78 (For convenience, the population is projected to 1968 from the available census values of 1967).

The relatively large survival ratios especially for males and the observation that some ratios are even higher than one, suggests that either there was some over-enumeration in 1978 in these ages and sex groups, or correspondingly there was some under-enumeration in 1967. Apparently, there was some under-enumeration of males in 1967 and it looks that the 1978 count might have been more complete.

The problem with these techniques, however, is that it is not always easy to determine whether the fault lies more with the first than with the second one. Additional consistency checks will have to be undertaken to determine the magnitude of the error in each.

The particular methods to be used to determine the accuracy of the census, as far as coverage is concerned, will depend a lot on the range of statistical data available in a particular situation, but in the final analysis, there is no substitute for a strict policy of rigid quality controls and supervision at all stages of an enumeration.

Table 3. COHORT SURVIVAL RATIOS AND OVERALL SURVIVAL RATIOS -- BY SEX, Tanzania (1968-78)

Cohort Survival Ratios

	Male			Female		
Age	Population aged x in 1968	Population aged x + 10 in 1978	Survival ratio	Population aged x in 1968	Population aged x + 10 in 1978	Survival ratio
0 - 4	1124298	1066645	.9497	1148918	1034802	.9008
5 - 9	1006237	841340	.8361	997628	877940	.8802
10-14	677203	586580	.8662	608920	747518	1.2196
15-19	528292	610325	1.1553	587936	703549	1.1955
20-24	389960	457537	1.1733	559473	504798	.9023
25-29	475323	439515	.9247	589389	446637	.7578
30-34	369548	321487	.8699	414003	348707	.8423
35-39	351506	320391	.9115	344908	311951	.9044
40-44	228698	233611	1.0215	243507	237052	.9735
45-49	260318	205252	.7885	238307	175181	.7351
50-54	183517	172414	.9395	189138	173357	.9166
55-59	112299	124810	1.1114	105737	116055	1.0976
60-64	113939	103046	.9044	120981	91085	.7528
65 +	375081	154933	.4131	329933	127169	.3836
All ages	6196227	5637886	.9099	6478496	5892801	.9096

.../50

Overall Survival Ratios (O.S.R.)

Age	Population aged x + in 1968	Population aged x + 10 in 1973	O.S.R.	Population aged x in 1968	Population aged x + 10 in 1978	O.S.R.
0 +	6196227	5637886	.9099	6478496	5982801	.9096
5 +	5071922	4571241	.9013	5329678	4857[illegible]	.9115
10 +	4065692	3729901	.9174	4332050	3980059	.9187
15 +	3388489	3143321	.9276	3723230	3337541	.8696
20 +	2860197	2532906	.8856	3135294	2533992	.8082
25 +	2470227	2075459	.8402	2575821	2029194	.7878
30 +	1994906	1635944	.8201	1966422	1582557	.7967
35 +	1625358	1314457	.8087	1572419	1233850	.7647
40 +	1273852	994066	.7804	1227511	921399	.7510
45 +	1045154	760455	.7276	984004	684847	.6960
50 +	784836	555203	.7074	745697	509666	.6835
55 +	601319	382789	.6366	556559	336309	.6043
60 +	429020	257979	.5275	450822	220254	.4886
65 +	375061	154933	.4131	329833	129169	.3916

4.4 CONTENT ERRORS EVALUATION

Any error in the characteristics that are reported is called 'content error'. Thus it can occur in any or all of the characteristics canvassed in an enquiry viz. age, sex, marital status, fertility, mortality, migration, education, economic activity, household and housing variables. An error in one characteristic can vitiate the value of other characteristics and some times the inclusion of a particular characteristic may be the cause for the error.

Such errors and biases can occur from erroneous or inconsistent reporting or recording, or failure to report or record information. Ignorance, misunderstanding or resistance of respondent and carelessness or misunderstanding on the part of the enumerator could result in errors in data. Clerical, or mechanical errors in editing, coding, data processing, imputation etc. could also vitiate data.

Content error could result from coverage error and vice versa. For example, in view of extra work involved in filling detailed questions for some persons of specific age-sex groups it may be possible that enumerators either change the characteristics (age - sex) of the respondent to avoid the extra work or just not do the extra work under one pretext or other. Yet a different type of coverage error affecting content is through omission or duplication of areas, households and members which could be selective. For instance, the usual omission of infants and very young children (of particular sex, sometimes) affects coverage and at the same time fouls up age-sex data. Another aspect is the problem of coverage of

young adult males, perhaps because of their mobility or other reasons (like head tax, military conscription) and of young females due to social and cultural reasons. The method of counting (de-facto, de-jure, time reference period for inclusion/exclusion of members) also might jeopardize coverage and could affect content.

Age is an important characteristics enquired into in all demographic surveys but due to one reason or the other in developing societies the reported ages are subject to various types of errors and biases. Even in developed societies biases have been noted in age data due to deliberate misstatement. Ignorance, apathy and the feeling that one's exact age is not very important results in several types of deficiencies in the reported age. One of the most important types of bias in age data is what is called 'digit preference error'. Ages are rounded to the nearest 5 or 10 and people are reported most as aged 10, 15, 20, 25 .... and rarely as 11, 19, 24, 27 etc.

From Table 4 giving single year of age data for mainland, it is clear that digits 0 and 5 are preferred and digits 1, 3 and 9 are avoided. This is more so among males than females, and more in rural than urban. In order to quantify the incidence of digit preference we present in Table 5 two indices - Whipples and Myers - by sex and residence. The conclusions drawn on the basis of the raw data are confirmed by the indices. The large value of the indices clearly point to the poor quality of single year of age data in the country in 1978.

Then there is the problem of age preference. If there be any reason for preferring to report or record a specific age then one finds that large numbers are reported in those ages. For example, if age at entry into school is 6 years, then a large number of children from above and below that age are reported in that age. If age 18 is considered as age for attaining majority and be eligible for benefits or rights (voting, marriage etc.) then preference for this age may result. This may be true of age which divides the young from the old say, age 30 or 40. The use of historical event calendars and other references might result in ages being reported around those events. Again, preference for digits could result from reported years of birth, age next birth day etc. It could also be affected by instructions to enumerators. Sometimes, errors in age reported in documents (which might have been estimated or rounded up) might result in age errors, if these documents are used by the enumerator. However, it is advantageous to have such documents, especially for young children (registration of births, identity cards) because at least further errors do not occur in age estimation. Age shifting is yet another phenomenon. Prestige, estimation of age based on marital or fertility status or physical appearance etc. could result in persons being pushed above or below their actual ages. The problem of 'not stated' ages is another. Also in some cases ages may be reported in broad categories like, infants, children, adults and old persons. In many societies age grade cohorts are well recognized and it may be possible only to allocate persons to such broad categories rather than to single year of ages.

.../52

Sex is another important characteristics collected in all demographic enquiries. It has a central role not only in study of demographic phenomenon (fertility, mortality, migration and population growth), but also most of the socio-economic activities and characteristics (marriage, household formation, education, economic activity etc.) are related to the sex of the population.

Even though it is not difficult to distinguish between the 2 sexes, still the reporting or coverage may differ between sexes due to one reason or other.

Table 4. POPULATION BY AGE IN SINGLE YEARS AND SEX

(Mainland - Tanzania)

AGE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
	TOTAL		RURAL		URBAN	
All Ages	8,350,492	8,686,006	7,175,951	7,602,596	1,174,511	1,083,410
0	294,731	314,618	255,373	273,496	39,358	41,122
1	282,420	290,797	245,820	253,895	36,600	36,902
2	297,668	310,751	259,781	270,938	37,887	39,813
3	322,327	329,841	284,096	291,548	38,231	38,293
4	316,146	382,483	280,684	291,372	35,462	37,111
5	312,907	308,868	278,109	273,373	34,798	35,495
6	291,415	296,106	260,032	262,858	31,383	33,248
7	262,383	267,111	231,823	234,713	30,560	32,398
8	266,482	279,380	237,440	248,107	29,042	31,273
9	215,549	216,874	192,534	191,665	23,015	25,209
10	243,707	241,532	218,240	214,105	25,467	27,427
11	165,427	168,016	146,917	147,607	18,510	20,409
12	239,120	221,631	215,990	196,290	23,130	25,341
13	191,826	184,410	171,324	162,210	20,502	22,200
14	197,838	192,965	177,392	170,617	20,446	22,348
15	189,125	176,427	165,966	154,580	23,159	21,847
16	168,900	161,080	146,553	138,905	22,347	22,175
17	147,287	146,278	126,156	124,616	21,131	21,662
18	200,581,	232,439	167,078	196,349	33,503	36,090
19	112,726	137,534	89,369	112,594	23,357	24,940
20	155,134	223,219	124,817	188,129	30,317	35,090
21	95,856	110,896	74,495	89,739	21,361	21,157
22	117,798	155,657	89,213	128,440	28,585	27,217
23	89,599	108,330	67,552	87,393	22,047	20,937
24	113,276	126,159	85,261	102,352	28,017	23,807
25	153,395	198,294	116,752	166,073	36,643	32,221
26	109,809	122,184	84,073	101,429	25,736	20,755
27	93,500	97,419	71,914	81,719	21,586	15,700
28	140,903	172,713	108,885	146,405	32,018	26,308
29	98,095	96,923	77,589	82,354	20,506	14,569
30	164,634	212,425	129,552	133,386	35,082	29,039

.../53

AGE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
	TOTAL		RURAL		URBAN	
31	67,323	58,906	54,237	51,198	13,086	7,708
32	94,232	100,894	74,324	88,421	19,908	12,473
33	63,699	58,069	51,440	50,274	12,259	7,795
34	56,362	61,344	45,256	53,872	11,106	7,472
35	130,173	145,322	105,580	126,553	24,593	18,769
36	82,481	82,021	67,550	72,853	14,931	9,168
37	55,529	49,009	45,904	43,023	9,625	5,986
38	99,873	105,669	83,072	93,724	16,801	11,945
39	61,474	55,408	51,103	48,943	10,371	6,465
40	122,333	160,109	101,759	142,250	20,574	17,859
41	38,348	33,990	32,393	30,979	5,955	3,011
42	69,643	66,280	58,494	60,025	11,149	6,255
43	43,307	42,581	36,915	38,999	6,392	3,582
44	38,487	35,356	33,547	32,350	4,940	3,006
45	115,456	118,076	98,302	105,513	17,154	12,563
46	48,309	44,111	42,163	40,565	6,146	3,546
47	37,365	32,149	32,972	29,534	4,393	2,615
48	74,606	78,538	65,728	71,861	8,878	6,677
49	37,103	33,088	32,300	30,129	4,803	2,959
50	96,242	122,635	84,514	110,819	11,728	11,816
51	25,547	19,869	23,030	18,250	2,517	1,619
52	43,229	36,971	38,373	34,106	4,856	2,865
53	30,008	21,764	27,405	20,156	2,603	1,608
54	31,099	28,185	28,334	26,354	2,765	1,831
55	61,564	59,474	53,826	53,345	7,738	6,129
56	43,626	31,326	39,463	29,008	4,163	2,228
57	24,226	16,177	22,099	15,051	2,127	1,126
58	48,174	42,829	44,482	39,991	3,692	2,838
59	23,267	22,452	21,385	21,118	1,882	1,334
60	83,617	91,405	76,182	82,723	7,435	8,682
61	14,946	13,919	13,910	13,017	1,036	902
62	21,580	22,672	20,082	21,173	1,498	1,499
63	15,089	14,700	14,005	13,811	1,084	889
64	30,652	26,538	28,408	24,667	2,244	1,921
65	46,360	50,978	42,343	46,166	4,017	4,812
66	16,085	15,016	14,954	14,066	1,131	950
67	17,059	11,731	15,908	10,988	1,151	743
68	29,120	25,341	27,203	23,189	1,917	2,152
69	13,067	10,880	12,362	10,288	705	592
70	50,156	50,124	46,484	45,969	3,672	4,155
71	9,395	8,051	8,764	7,548	631	503
72	13,880	11,666	12,963	10,787	917	879
73	15,484	11,043	14,714	10,299	770	744
74	10,304	6,909	9,748	6,405	556	504
75	23,805	21,735	22,096	20,102	1,709	1,633

.../54

Table (Contd..)

AGE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
	TOTAL		RURAL		URBAN	
76	9,437	8,326	8,799	7,809	638	517
77	6,518	4,986	6,097	4,688	421	298
78	18,840	14,544	17,793	13,640	1,047	904
79	5,832	4,173	5,460	3,933	372	240
80	20,877	21,664	19,592	20,014	1,285	1,650
81	3,462	2,535	3,351	2,354	111	181
82	5,062	4,164	4,815	3,995	247	169
83	3,016	2,496	2,881	2,394	135	102
84	3,081	1,891	2,925	1,775	156	116
85	7,778	6,747	7,382	6,306	396	441
86	2,853	1,668	2,747	1,554	106	114
87	1,968	2,007	1,868	1,946	100	61
88	4,000	2,632	3,777	2,480	223	152
89	3,965	3,633	3,752	3,449	213	184
90+	29,555	19,913	27,888	18,541	1,667	1,372

Table 5. DIGIT PREFERENCE INDICES (WHIPPLES AND MYERS) BY SEX-RURAL, URBAN - Mainland

Whipples Index

RURAL		URBAN	
MALE	FEMALE	MALE	FEMALE
163.2	183.3	161.5	181.5

Blended Percentage and Myers Index

Digit		0	1	2	3	4	5	6	7	8	9	Index
Rural	M	15.2	6.5	9.7	7.5	8.5	14.1	9.8	7.9	13.1	7.8	24.8
	F	18.4	5.9	9.8	6.9	8.0	14.3	8.8	6.9	13.5	7.5	32.4
Urban	M	15.0	6.7	10.2	7.3	8.2	14.2	9.4	7.7	12.9	8.5	24.6
	F	17.3	6.4	9.6	7.2	8.0	14.1	8.6	7.1	13.4	8.4	29.6

.../55

For example, due to fear of evil eye, prestige or value of children, the reporting of children of a particular sex may be deficient or defective. Instances of sex being wrongly reported also have been noted. Omission or duplication of persons of one sex may be more or less due to method of enumeration, coverage, mobility, cultural practices etc. Also the reporting of ages differ by sex. In most cases, ages are reported by male members of households (heads or most knowledgeable persons) and the estimation of female ages may be more in error. This becomes more so when the enumerator may have no opportunity to see the persons concerned. The tendency of over-estimating ages of females when they are married and have several children especially when they are in their early teens and under-estimating ages at older ages due to fertility status (to be included in reproductive ages) have been noted. For males, the problem may be more in respect of over-estimation of age due to prestige for old age in the society.

Sometimes, enumerators may not enter sex of members because this can be inferred from the names. But in societies when names are same or similar, imputation of sex in the office can result in errors. Coding also may result in further errors if code 1 is used for males and 2 for females and the writing of codes is not legible and possibility for mix up can occur. (e.g. 1 can look like 2, if not properly written). It is advantageous to use codes 1 for male and 3 for female.

Even though single year of age-sex data seem to be riddled with a large amount of error, perhaps by grouping the data, some of these errors may be removed or at least reduced. Table 6 presents data by five year age groups. Even though the smoothness has certainly improved over the single year values, still some undulations and fluctuations are obvious. One measure to study the fluctuations is the United Nations score method. In this, separate age ratios by sex and sex ratios by age are calculated and in normal circumstances, these ratios should not deviate much from each other and from 100.

Table 7 and 8 show the results for 1967 and 1978. It is clear that there is an improvement over time, female data is poorer, and mainland is better than Zanzibar. According to the UN, if the score is less than 20, then the data may be considered acceptable. Beyond 40, the data are classified as very inaccurate. Since the values are much higher than 40 in all cases, it can be concluded that the quality of the data leaves much to be desired. From Table 9 and 10, we note that the quality of data in Tanzania is not much different from those of neighbouring countries. The fluctuations of age ratios and the specific pattern of sex ratios clearly indicate the age preferences; age shifting and other reporting errors. Exaggeration of ages is also evident from the unusually large proportion of older persons, especially among males and this is confirmed also by the survival ratios in Table 3.

.../56

Table 6. POPULATION BY SEX AND FIVE YEAR AGE GROUPS

Tanzania

AGE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
	TOTAL		RURAL		URBAN	
TOTAL	8,587,086	8,925,525	7,337,247	7,762,462	1,249,839	1,163,063
0 - 4	1,557,979	1,620,439	1,356,977	1,413,054	201,002	207,385
5 - 9	1,391,221	1,412,285	1,230,559	1,241,430	160,662	170,855
10-14	1,066,645	1,034,802	949,775	907,407	116,870	127,395
15-19	841,340	877,940	710,856	742,298	130,484	135,642
20-24	586,580	742,518	450,409	607,677	136,171	134,841
25-29	610,325	703,549	468,225	588,069	142,100	115,480
30-34	457,537	504,798	361,905	436,116	95,632	68,582
35-39	439,515	446,637	359,620	391,070	79,895	55,567
40-44	321,487	348,707	269,213	311,719	52,274	36,988
45-49	320,391	311,951	276,220	281,506	44,171	30,445
50-54	233,611	237,052	206,735	214,866	26,876	22,186
55-59	205,252	175,181	184,048	160,431	21,204	14,750
60-64	172,414	173,357	157,280	159,447	15,134	15,910
65-69	124,310	116,055	114,883	106,012	9,927	10,043
70-74	103,046	91,085	95,495	83,396	7,551	7,689
75-79	66,150	54,700	61,532	50,812	4,618	3,888
80 +	88,783	72,469	83,515	67,152	5,268	5,317

Mainland

AGE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
TOTAL	8,350,492	8,686,006	7,175,981	7,602,596	1,174,511	1,083,410
0 - 4	1,513,292	1,574,490	1,325,754	1,381,249	187,538	193,241
5 - 9	1,348,735	1,368,340	1,199,938	1,210,717	148,797	157,623
10-14	1,037,919	1,008,554	929,864	890,829	108,055	117,725
15-19	818,619	853,757	695,122	727,044	123,497	126,713
20-24	571,666	724,260	441,339	596,082	130,327	128,208
25-29	595,702	687,532	459,213	577,980	136,489	109,552
30-34	446,249	491,639	354,808	427,151	91,441	64,488
35-39	429,528	437,429	353,208	385,097	76,520	52,332
40-44	312,119	338,317	263,108	304,603	49,011	33,714
45-49	312,839	305,961	271,465	277,602	41,374	28,359
50-54	226,124	229,426	201,655	209,686	24,469	19,740
55-59	200,858	172,168	181,256	158,513	19,602	13,655
60-64	165,885	169,285	152,587	155,391	13,298	13,894
65-69	121,691	113,946	112,770	104,697	8,921	9,249
70-74	99,217	87,792	92,672	81,008	6,545	6,784
75-79	64,431	53,763	60,245	50,171	4,186	3,592
80 +	85,618	69,347	80,977	64,806	4,641	4,541

.../57

Table 6 (Contd...)

Zanzibar

AGE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
	TOTAL		RURAL		URBAN	
TOTAL	236,592	239,519	161,266	159,865	75,326	79,654
0 - 4	44,687	45,949	31,223	31,805	13,464	14,144
5 - 9	42,486	43,945	30,621	30,713	11,865	13,232
10-14	28,726	26,248	19,911	16,578	8,815	9,670
15-19	22,721	24,182	15,734	15,254	6,987	8,928
20-24	14,914	18,259	9,070	11,625	5,844	6,634
25-29	14,623	16,017	9,012	10,089	5,611	5,928
30-34	11,283	13,159	7,097	8,965	4,191	4,194
35-39	9,987	9,209	6,412	5,974	3,575	3,235
40-44	9,369	10,389	6,105	7,115	3,263	3,274
45-49	7,551	5,990	4,755	3,904	2,796	2,086
50-54	7,488	7,626	5,080	5,180	2,408	2,446
55-59	4,394	3,013	2,792	1,918	1,602	1,095
60-64	6,529	6,072	4,693	4,056	1,836	2,016
65-69	3,118	2,109	2,113	1,315	1,005	794
70-74	3,828	3,294	2,823	2,388	1,005	906
75-79	1,719	937	1,287	641	432	296
80+	3,165	3,121	2,538	2,345	627	776

Table 7. AGE RATIOS, SEX RATIOS AND JOINT SCORES, Tanzania, Mainland and Zanzibar

Tanzania

	Sex ratio		Age ratio			
AGE	1967	1978	1967	1978	1967	1978
0 - 4	98	96				
5 - 9	101	99	112	106	114	106
10-14	111	103	88	96	77	90
15-19	90	96	99	102	101	99
20-24	70	79	78	81	95	94
25-29	81	87	125	117	121	113
30-34	89	91	89	87	89	88
35-39	102	98	118	113	105	105
40-44	94	92	75	85	83	92
45-49	94	103	126	115	110	107
50-54	109	99	99	89	110	97
55-59	97	117	76	101	68	85
60-64	106	100	119	105	133	119

.../58

Table 7 (Contd...)

Tanzania

Age	Sex ratio		Age ratio			
	1967	1978	1967	1978	1967	1978
65-69	94	108	90	91	81	98
All ages	95	96				
Scores	10.9	8.8	15.8	10.0	15.5	9.0

Mainland

Age	Sex ratio 1967	Sex ratio 1978	Age ratio 1967	Age ratio 1978	Age ratio 1967	Age ratio 1978
0 - 4	98	96				
5 - 9	101	99	111	106	113	106
10-14	111	103	89	96	77	91
15-19	90	96	99	102	101	99
20-24	70	79	78	81	95	94
25-29	80	87	125	117	122	113
30-34	89	91	89	87	88	87
35-39	101	98	118	113	106	105
40-44	94	92	74	84	82	91
45-49	108	102	128	116	112	108
50-54	97	99	97	88	108	96
55-59	105	117	77	103	70	86
60-64	103	98	116	103	130	118
65-69	103	107	92	82	84	89
All ages	95	96				
Score	11.3	8.8	15.6	10.2	15.1	9.0

Zanzibar

Age	Sex ratio 1967	Sex ratio 1978	Age ratio 1967	Age ratio 1978	Age ratio 1967	Age ratio 1978
0 - 4	98	97				
5 - 9	97	97	129	116	139	122
10-14	123	109	70	88	55	77
15-19	94	94	92	104	90	109
20-24	67	82	80	80	109	91
25-29	89	91	120	112	106	102
30-34	91	86	106	92	119	104
35-39	126	108	93	97	69	78
40-44	98	90	112	107	152	137
45-49	145	126	77	90	53	67
50-54	103	98	150	125	221	169
55-59	165	146	50	63	33	44
60-64	119	107	223	174	323	237
65-69	135	148	49	60	31	45
All ages	101	101				
Score	33.3	21.8	32.9	20.6	56.8	36.8

Table 8. U.N. JOINT SCORES 1967 and 1978 - Tanzania, Mainland and Zanzibar

	Tanzania		Mainland		Zanzibar	
	1967	1978	1967	1978	1967	1978
U.N. Joint Score	64.0	45.4	64.6	45.6	189.6	122.8

Table 9. AGE RATIOS FOR TANZANIA COMPARED WITH THOSE OF SOME NEIGHBOURING COUNTRIES

Age		Kenya (1969)	Tanzania (1967)	Tanzania (1978)	Uganda (1969)	Zambia (1969)
5 - 9	M	103	112	106	99	104
	F	104	113	106	101	106
10-14	M	97	109	96	100	96
	F	92	77	90	91	88
15-19	M	98	99	102	91	94
	F	98	101	99	92	90
20-24	M	94	78	81	88	89
	F	94	95	94	96	110
24-29	M	99	125	117	109	101
	F	110	121	113	112	97
30-34	M	93	89	87	100	96
	F	89	69	88	97	102
35-39	M	106	117	113	103	115
	F	106	105	105	97	104
40-44	M	91	75	85	93	86
	F	94	83	92	99	88
45-49	M	106	126	115	98	116
	F	96	110	107	89	109
50-54	M	92	98	89	111	80
	F	105	110	97	125	92
55-59	M	98	76	101	79	142
	F	88	68	85	69	103
60-64	M	108	119	105	124	64
	F	114	133	119	139	83

.../60

Table 10. SEX RATIOS BY AGE FOR TANZANIA COMPARED WITH THOSE OF SOME NEIGHBOURING COUNTRIES

Age	Kenya (1969)	Tanzania (1967)	Tanzania (1978)	Uganda (1969)	Zambia (1969)
0 - 4	101	98	96	98	97
5 - 9	103	101	99	100	99
10-14	106	111	103	110	107
15-19	103	90	96	100	94
20-24	95	70	79	89	70
25-29	85	81	87	93	78
30-34	94	89	91	102	81
35-39	95	102	98	109	99
40-44	96	94	92	104	102
45-49	105	109	103	113	112

A look at the sex ratio at young ages indicates some relative omission of male more than female children. The percentage distribution given in Table 11 indicates fluctuations and also the possibility of relative under-enumeration at age 0 to 4. One method to check coverage at young ages is by the method of reverse survival. The reversed populations based on the enumerated children aged 0 to 4 in 1978 for the period of five years before the census would give the births of 1973 to 1978 from which one could derive birth rates and compare with expected birth rate and sex ratio at birth. Using North model life tables level 11 and growth rate of 2.7% during 1973 to 1978, we estimate an annual average of 384400 male and 389000 female births resulting in a sex ratio at birth of 99 and crude birth rates of 48 for male and 46 for females. The estimated crude birth rate is 49 and sex ratio at birth more than 100. On both counts, it looks that some children have been omitted.

In regard to marital status, the confusion between 'single' and 'never married' can result in errors. Also there is possibility of reporting of 'divorced' or 'separated' persons as 'married' and also some 'widows' or 'widowers' as 'married'. Again age estimation may be based on marital status and hence can affect both chracteristics.

In respect of fertility information, in addition to age errors, the problem of under reporting of births, memory lapse, vagueness about time reference period etc. might vitiate the data. There could be selectivity in reporting of births by sex (in one case, it was noticed that since fertility question was meant only for female population, information also were only collected on female births). It is advantageous if data were collected by sex of children which then could be used for evaluation of the data.

Table 11. PERCENTAGE AGE DISTRIBUTION BY SEX, Tanzania, Mainland and Zanzibar, 1978

AGE GROUP	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
	TOTAL		RURAL		URBAN	
			Tanzania			
	%	%	%	%	%	%
0 - 4	18.14	18.16	18.49	18.20	16.08	17.83
5 - 9	16.20	15.82	16.77	15.99	12.85	14.69
10-14	12.42	11.59	12.94	11.69	9.35	10.95
15-19	9.80	9.84	9.69	9.56	10.44	11.66
20-24	6.83	8.32	6.14	7.83	10.90	11.59
24-29	7.11	7.88	6.38	7.58	11.37	9.93
30-34	5.33	5.66	4.93	5.62	7.65	5.90
35-39	5.12	5.00	4.90	5.03	6.39	4.78
40-44	3.74	3.91	3.67	4.02	4.18	3.18
45-49	3.73	3.50	3.76	3.63	3.53	2.62
50-54	2.72	2.66	2.82	2.77	2.15	1.91
55-59	2.39	1.96	2.51	2.07	1.70	1.27
60-64	2.01	1.94	2.14	2.05	1.21	1.37
65-69	1.45	1.30	1.57	1.37	.79	.86
70-74	1.20	1.02	1.30	1.07	.60	.66
75-79	.79	.61	.84	.65	.17	.33
80 +	1.03	.81	1.14	.87	.42	.46
			Mainland			
0 - 4	18.12	18.13	18.47	18.17	15.97	17.84
5 - 9	16.15	15.75	16.72	15.93	12.67	14.55
10-14	12.43	11.61	12.96	11.72	9.20	10.87
15-19	9.80	9.83	9.69	9.56	10.51	11.70
20-24	6.85	8.34	6.15	7.84	11.10	11.83
25-29	7.13	7.92	6.40	7.60	11.62	10.11
30-34	5.34	5.66	4.94	5.62	7.79	5.95
35-39	5.14	5.04	4.92	5.07	6.52	4.83
40-44	3.74	3.89	3.67	4.01	4.17	3.11
45-49	3.75	3.52	3.78	3.65	3.52	2.62
50-54	2.71	2.64	2.81	2.76	2.08	1.82
55-59	2.41	1.98	2.53	2.08	1.67	1.26
60-64	1.99	1.95	2.13	2.04	1.13	1.28
65-69	1.46	1.31	1.57	1.38	.75	.85
70-74	1.19	1.01	1.29	1.07	.56	.63
75-79	.77	.62	.84	.66	.36	.33
80 +	1.03	.80	1.13	.85	.39	.42
			Zanzibar			
0 - 4	18.89	19.18	19.36	19.89	17.87	17.76
5 - 9	17.96	18.35	18.99	19.21	15.75	16.61

.../62

AGE GROUP	TOTAL MALE	TOTAL FEMALE	RURAL MALE	RURAL FEMALE	URBAN MALE	URBAN FEMALE
10-14	12.14	10.96	12.35	10.37	11.70	12.14
15-19	9.60	10.10	9.76	9.54	9.28	11.21
20-24	6.30	7.62	5.62	7.27	7.76	8.33
25-29	6.18	6.69	5.59	6.31	7.45	7.44
30-34	4.77	5.49	4.40	5.61	5.56	5.27
35-39	4.22	3.84	3.98	3.74	4.75	4.06
40-44	3.96	4.34	3.79	4.45	4.33	4.11
45-49	3.19	2.50	2.95	2.44	3.71	2.62
50-54	3.16	3.18	3.15	2.24	3.20	3.07
55-59	1.86	1.26	1.73	1.20	2.13	1.37
60-64	2.76	2.54	2.91	2.54	2.44	2.53
65-69	1.32	.88	1.31	.82	1.33	1.00
70-74	1.62	1.38	1.75	1.49	1.33	1.14
75-79	.73	.39	.80	.40	.57	.37
80 +	1.34	1.30	1.57	1.47	.83	.97

Misunderstanding of the question also can affect data. For example, report on children ever born alive may exclude children who died (especially if it occured in early childhood and a long time ago). Instances have been noted where children who are now grown up but still living may also be omitted because they are no longer 'children'. Question on children born in households during past year may be affected by the fact that such births occur to members who may not be present during the survey or may not even be usual members of households. If questions were directed only to those members who are usual members at the time of survey, then births to persons who died or moved out may be left out. In any case, the experience has been that a large number of events are not reported in such enquiries and indirect methods have to be used to estimate the level of fertility, if not the pattern.

This is more so in respect of information on deaths. Quality of information on deaths in households during past year has been very poor. One reason for breakup of households could be a death and in any case, death is a sad event and no one likes to be reminded of it and the earlier it is forgotten, the better. As mentioned earlier, reporting on number of children dead for mothers also might be affected by selective omission. Sometimes question is asked on year of birth of last child and its survival status. In such cases, the question of year of birth may be affected by memory lapse and difficulty for people to identify the year to which the birth of the last child belongs. It is sometimes noted that many events of past 2 or 3 years are reported as having occured during last year. There is also the possibility of omission of a child who died immediately after birth.

.../63

There are some questions which are included in questionnaires to arrive at estimates of mortality viz those on parent, spouse or sibling survival. In addition to age reporting errors, there is also the problem of foster parents, remarriage, selective survival etc. affecting the information.

Some of the methods also need information on mean age at birth of children (for females and males) which again may be difficult or outright unacceptable.

In study of migration the method of enumeration the coverage and the question to distinguish the mover from the non mover may all contribute to biases. In many societies, the place of birth may be where the mother went for delivery which may not have much to do with later movements. Place of previous residence, (at a specified time or otherwise) may also not bring out a true picture of movements. The question on the rural/urban nature of place of birth or previous residence etc may be confusing when the status of the place may have changed since the movement. Changes in names of places, area changes etc also may result in problem in identification of locations. There is also the tendency for people to report the nearest town as place of birth or previous residence even though the actual places may be the village around the town.

The question of duration of residence should also be handled carefully. Instances of non movers being given their age as duration of residence in the place of enumeration are not rate. Rounding problems also are observed in reporting of duration. Problems in editing, coding etc also may result in further deterioration of data as e.g., when no distinction is made between persons moving from urban area to another urban area within a district (urban to urban intra district migrant) from those moving within the same urban area in the district. This is true of intra district rural to rural migrants also.

When it comes to international migration, the problem may be more difficult where tendency for wrong reporting of nationality, citizenship and place of birth may be higher. This may be due to political, phsychological, geographical or historical reasons. Sometimes, it may even result in under-enumeration.

In this connection it may be worth while to consider carefully the question of collection of data on ethnic, tribal, language, religion and other such characteristics. In cases where such questions may create fear or suspicion in the minds of the population, it may be advisable not to collect these information in a survey even though these information may have very great value to the sociologist, anthropologiest, planner etc. Instances where the inclusion of questions on such aspects have led to poor enumerations are many.

In regard to information on educational characteristics there is the problem of status climbing by some and reporting educational attainment higher than actual for prestige. Sometimes it could be due to ignorance when a child who has appeared for an examination but whose result

.../64

is not announced, is reported as having attained the level of the examination. There is also the confusion in some cases as to whether the level attained is the one he has completed or the one he is attending. The problem of distinguishing between the literate and illiterate is yet another. There is also the problem of educational statistics affecting age reporting whereby the age of a child is estimated by his level of education he is attending. Instances where a child is assigned a date of birth (date, month and year) on enrolment into a school are well known. In one country, it is noted that a large proportion of children who have gone through school are found to be reported as born on 1st of January of various years and this is due to assignment of 1 January as date of birth at time of enrolment. On the whole however, data from educated respondents are found to be better than those from the uneducated excepting those items where deliberate misstatements play an important role.

Coming to information on economic characteristics, there could be genuine problems in definitions, concepts etc. Status climbing is noted in respect of economic status, occupation etc.

Since, in many cases, information on marital status, fertility, mortality, migration, education and economic activity etc. are only collected on special segments (age - sex) of the population, and the details could be time consuming, there has been noted to be a tendency for enumerators to alter ages of respondents to reduce the work load, especially when there is no special payment for the extra work. However, when payment is according to numbers of schedules canvassed, an opposite tendency to include ineligible persons has also been noted, resulting in errors in age and economic characteristics.

One of the observations on data on households has been that there is a tendency to under represent one person households, especially in urban areas. The selective coverage of persons within households may alter house hold composition and structure. Another observation has been that when questionnaires have a fixed number of lines for members of households (say 10) then either due to inertia on the part of the enumerator to continue with enumeration of members of large households (with more than 10 members) or due to improper processing, a disproportionate number of households tend to be reported with members equal to number of lines (10 say). One dodge has been to increase the number of lines in the schedule and the other is strict supervision to ensure that continuation sheets are used for large households and marked and identified as such.

Very little evaluation has been carried out regarding quality of information on housing characteristics. Firstly, the problem is lack of knowledge and information and secondly lack of interest of demographers in such purely non demographic characteristics. Even it has been questioned whether a population census is an appropriate forum to collect information on housing. However, having collected the information, it is essential that they be evaluated, analyzed and utilized. This is true of information on physical status - which a few countries have collected in their censuses. But housing characteristics will form an important aspect of household surveys and there is need for evolution of techniques for evaluating, analyzing interpreting and utilizing them.

.../65

5. DATA PRESENTATION

Even though tabular presentation of data is an essential pre-requisite of the analysis of the information, some intermediate stages of analysis may be needed to arrive at ideas about the data, the information portrayed and the types of evaluation and analyses to be attempted.

Graphic presentation of data is an extremely useful and flexible medium for explaining, interpreting and analysing numerical facts by means of points, lines, areas and other geometric forms and symbols. They make possible the presentation of quantitative data in a simple, lucid and effective manner and facilitate comparison of values, trends and relationships. Moreover, graphical representation are more effective in creating interest and in appealing to the attention of the reader, they can be grasped more easily and with clarity, they are comprehensive and better balanced than tabular presentation and cannot only save space and time, they are also more revealing and stimulating and aid in analysis and interpretation.

5.1 RECTILINEAR CO-ORDINATE CHARTS

Arithmetic line chart

The arithmetic line chart is one of the several types of rectilinear co-ordinate graphs and is obtained by plotting one or more series of figures on a co-ordinate surface in which the successive points are joined together in the form of a curve. It is particularly effective in portraying time series such as movements or trends over a period of years or variations covering shorter periods like months, days, hours etc.

Figures 6 and 7 respectively give the sex ratios by single year of age for Tanzania mainland as obtained from the 1978 census and percent distribution of the mainland population by age and sex for 1967 and 1978 by sex.

Bar chart

The bar chart is another most useful and simple techniques in graphic presentation. It is particularly appropriate for comparing the magnitude or size of co-ordinate items or of parts of a total. The length of each bar or of its components is proportional to the quantity or amount of each category represented. The bars are arranged in order of size starting with the largest and they are presented horizontally. Sometimes, the bars are arranged alphabetically, geographically or other systematic ordering of data. The width of the bars should be uniform and the space between bars should be around half the width of the bars and the proportion of the length of bars and the total height of the chart should be maintained for aesthetic purposes.

Sometimes, a single bar may be subdivided into component categories by appropriate shading or hatching. A bar chart with bars presented

FIGURE 6 SEX RATIOS FOR THE MAINLAND: 1967 AND 1978

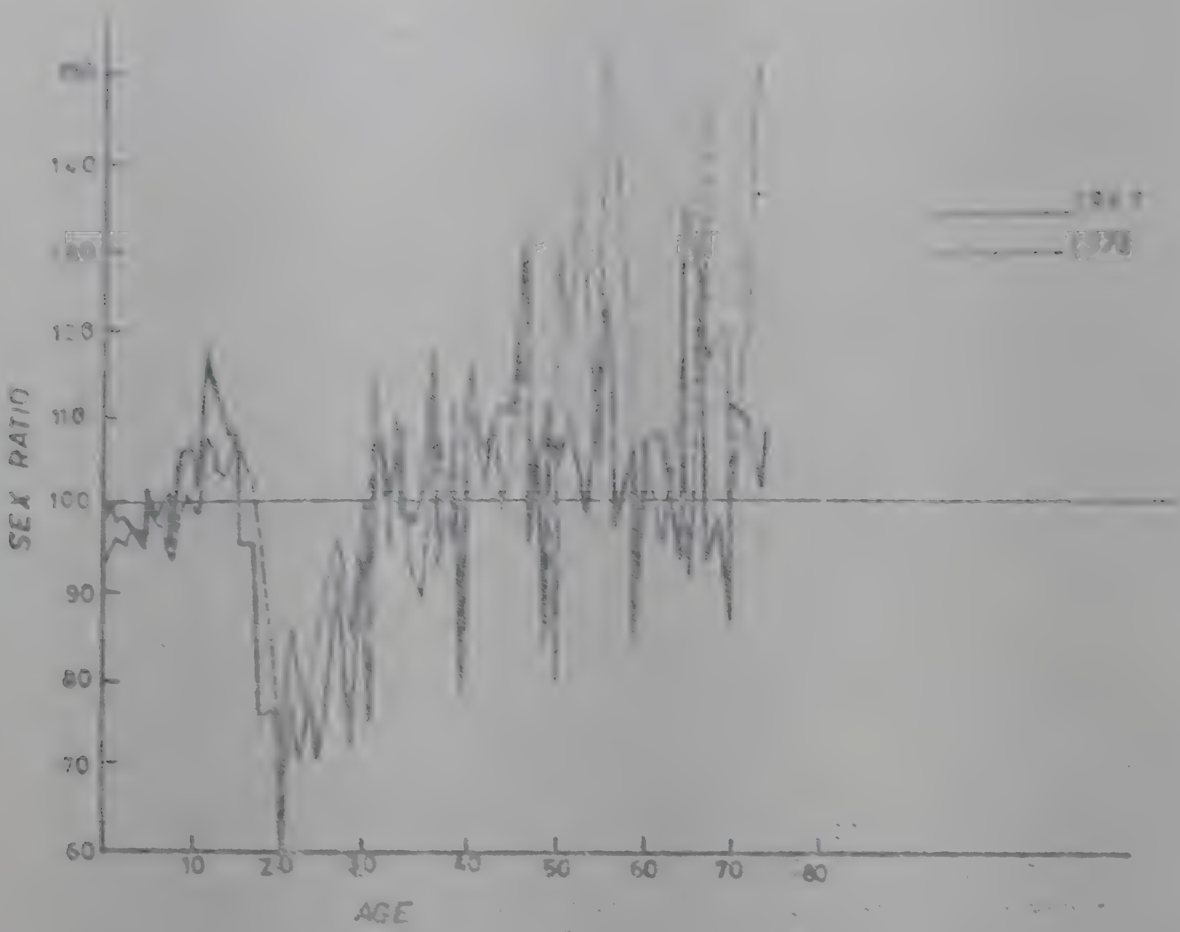
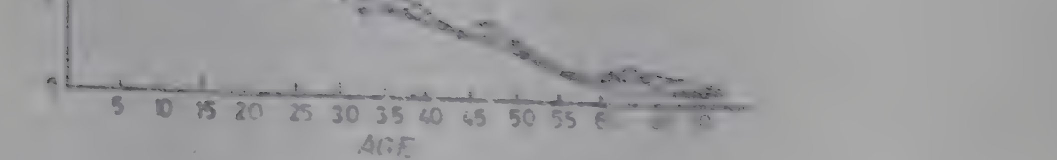


FIGURE 7 PERCENT DISTRIBUTION OF THE MAINLAND POPULATION BY AGE AND SEX: 1967 AND 1978



vertically is called a column chart and this type of chart is valuable in portraying time series, especially if the number of plotted values is not very large.

Histogram or frequency histogram

A histogram consists of a set of rectangles having bases on a horizontal axis with centres at the class marks and widths equal to the class interval sizes and areas proportional to class frequencies, i.e. a class having double the width will have only half the height for a given frequency as compared with a class with normal width.

Pie chart

A pie chart is yet another graphic method of presenting data and sometimes more vivid than the subdivided or segmented or component bar chart. A circle of appropriate radius is drawn and segments marked proportional to frequencies or amounts of the categories. Either the degrees portrayed by segments could be marked or it is possible to use protractors marked in percentages rather than in degrees. It is aesthetic to arrange categories according to relative size and the segments may be marked by proper hatching or shading. Labels should as far as feasible be put inside the segments or just outside. It is advisable not to have several categories represented- perhaps four to six may be ideal. Also care should be taken to distinguish even the smallest category appropriately.

Population pyramid

Population pyramid is a very effective and widely used method of graphically depicting the age-sex composition of a population. It is designed to give a vivid picture of the age-sex structure of a population, indicating either single ages, 5-year groups or other age combinations. The basic pyramid form consists of bars, representing age groups in ascending order from the lowest to the highest, placed one over the other resulting in the shape of a pyramid.

Usually bars for males are presented on the left of a central vertical axis and those for females on the right. The length of bars at any age will represent either the absolute number or the percentage to the total. Generally age is indicated in the central certical axis or either to the right, left or both depending on space, display needs etc. In general, the age groups in a given pyramid must have the same class interval and must be represented by bars of equal thickness. To minimize visual distortion, it is advisable to have the thickness of a bar such that the resulting pyramid has a base about 1.5 times the height. For comparative purposes, it is advisable to have the ratio of the thickness of a bar to the distance for a unit amount or 1% on the horizontal scale to be the same. Generally pyramids are drawn with five year age group data. However, usually five year of age group information is available only to age 80, 85 or 90. It may be sufficient for many practical purposes to truncate the pyramid at the highest five

.../67

FIGURE 8 .....POPULATION PYRAMID FOR TANZANIA MAINLAND AND FOR ZANZIBAR

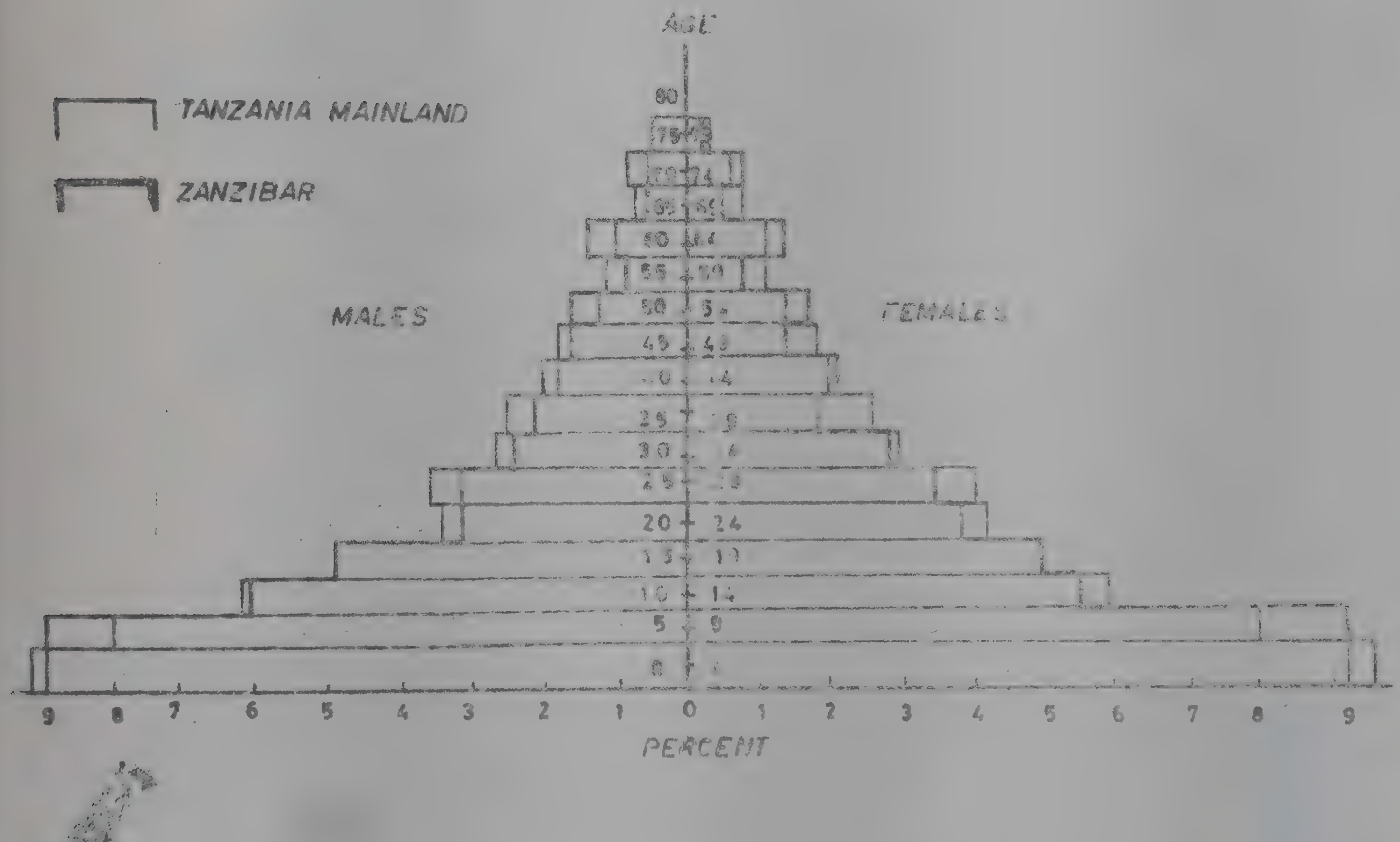


FIG. 9. RURAL POPULATION DENSITY BY DISTRICT 1978.

year age group resulting in most cases with 15 to 17 classes. In some countries, five year of age group data may be available only for certain young age groups and for older ages broader age groups like ten years may be presented. In such cases either it is preferable to prepare the pyramid for ten year age group data, or split the ten year group data into five year groups by using interpolation methods and present the resulting pyramids with five year groups.

In drawing the pyramid with percentages, it is necessary to calculate the percentages on the basis of the grand total for the population including the population in a terminal age group and for the two sexes together.

Figure 8 presents the population pyramid for Tanzania mainland and Zanzibar.

Density charts

A conventional way of representing population density in an area is by shading or hatching on the map. Greater densities are represented by darker shadings. It has to be kept in mind that since the densities are usually available for larger units of areas, the averaging process may hide some of the internal variations in densities. The smaller the areal unit for which the information is available, the better and more realistic will the representation be.

Figure 9 presents the density chart for rural population of Tanzania, 1978.

Dot maps

The commonest method of representing the distribution of the absolute number of inhabitants is a dot map, in which a small dot or spot of constant size represents a round number such as 100 or 1000 etc. For highlighting population concentration in specific areas within a given unit of area, it may be necessary to arrange the dots in a way that conveys the concentration and its size. For example, in maps of population distribution for a country or area containing both thinly settled rural territory and large urban agglomerations, dots of a different colour with a higher value may be advisable. Circles of varying sizes also may be utilized with area proportional to size of population, fixing the size for the largest place so that, as far as feasible, circles do not overlap and otherwise blur the picture. If they do overlap, different shadings may be used to distinguish them.

Flow maps

An example of charts in perspective projection is the flow map which is especially useful in the study of migration. In these maps, arrows of varying width indicate the volume and direction of migration indicating origin and destination. The width of each arrow is directly

.../68

proportional to the volume of migration and the length and position of the arrows identify the areas of origin and destination. The exact position and curvature of the curves can be modified to suggest the principal routes of the migrants. The arrow may reflect a division of the migrants from a given origin into several streams or a fusion of streams into a single current.

5.2 LOGARITHMIC AND SEMI LOG PAPERS AND LOGIT FUNCTIONS

In the presentation of data, in addition to use of ordinary graph paper (where the axis are marked in linear scale) sometimes we use graph papers where either one or both axes are in logarithmic scale. Such papers are called semi logarithmic (one axis in log scale) or double logarithmic (both axes in log scale) papers.

For example, in the linear model for population growth (also known as arithmetic growth model or simple interest law), the equation for population at a time t is given by:

$$P_t = P_o \ (1 + rt) \text{ where}$$

P_t = Pop. at time t, P_o = Pop. at base period, r = arithmetic growth rate, and t = no. of years since base period. The graph of this function with population on the y axis and time t on the x axis will be a straight line.

On the other hand, if population were growing by geometric law (compound interest law) then the equation will be:

$$P_t = P_o \ (1 + r)^t$$

In this case, if we take logarithm of both sides, we get

$$\log P_t = \log P_o + t \log (1 + r),$$

which has now a linear form. Hence under geometric law, if we plot $\log P_t$ on y axis and time t on x axis, we would find a straight line graph.

The same would be the case if growth of population were by exponential law (compound interest law calculated at every micro units of time) i.e.

$$P_t = P_o \ e^{rt} \text{ where } e = \text{ exponential function} = 2.7183.$$

Here we take logarithm either to base 10 or e and plot these values on y axis. Thus in the above 2 cases semi log papers would be useful and plot would indicate linear form. Thus when the interest is in study of amount of growth, the ordinary graph paper will be sufficient. But when we need the relative growth or rate of change, we use log scale and log papers.

If we consider the equation:

$y = a x^b$, then if we take log of both sides we get

$$\log y = \log a + b \log x$$

Under such a function, it we plot the points on a double log paper, we would find a straight line graph.

Paretos Curve

Pareto's curve which is generally used to represent income distribution has the form:

$$y = a x^{-b} \text{ where}$$

x = income size and y = no. of persons having that income or larger.

Another situation where a similar curve is applicable is in study of size of cities (y) and their ranks (x). For example, it has been observed that

$y = a x^{-b}$ where y = size of town and x is the rank of the town.
In both cases, a plot of y and x values in double log paper will produce a straight line graph.

Logit Function

One function derived from natural logarithms is the logit function defined as:

$$\text{logit } x = \tfrac{1}{2} \ln \left(\frac{1 - x}{x} \right) \text{where}$$

ln = natural logarith i.e. log to base e.

(Here x lies between o and 1).

This function is very much useful in the study of mortality.

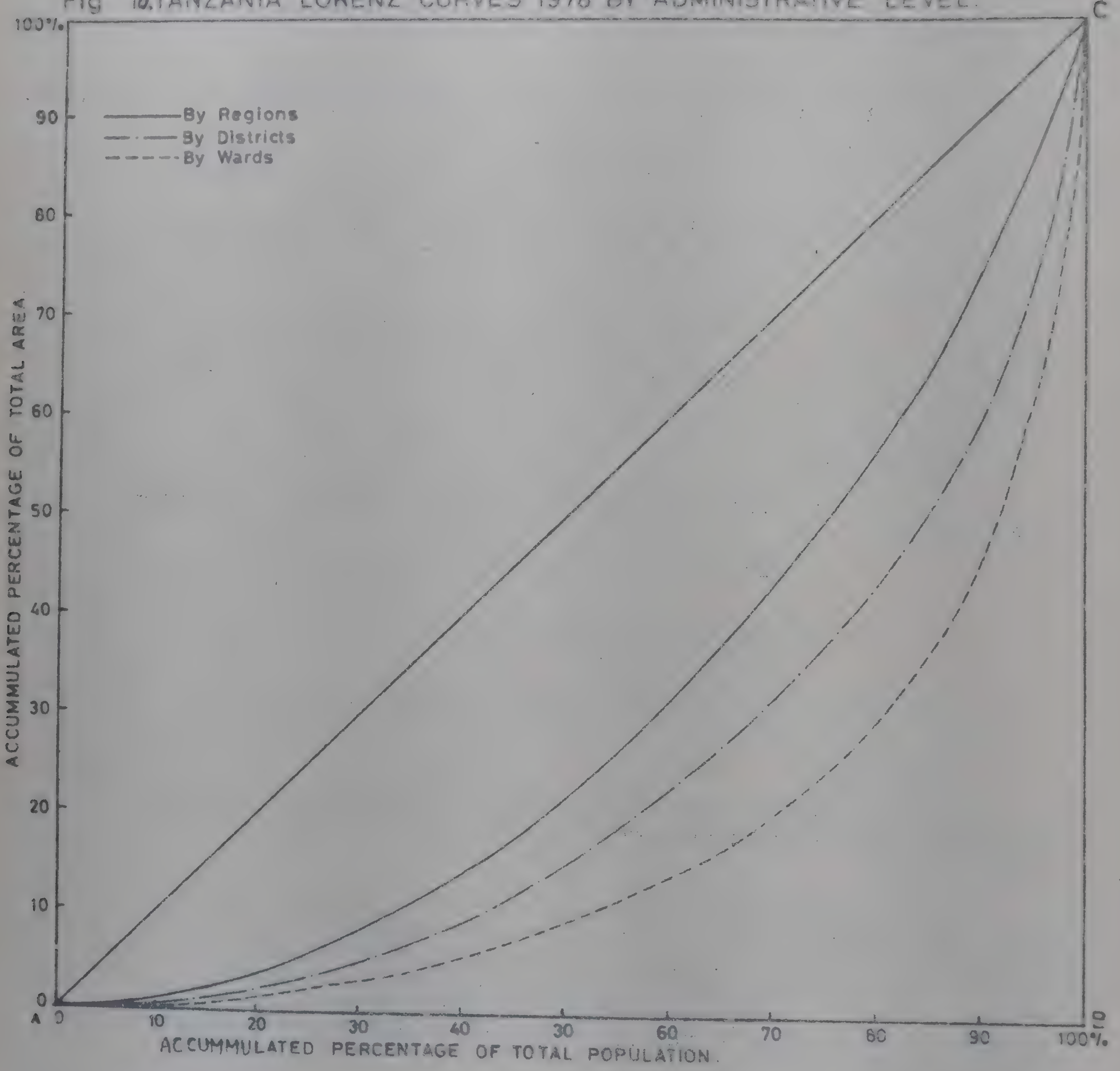
Lorenz curve

Another important curve which has been used to depict the state of concentration of population and of other demographic aggregates is the 'Lorenz curve'.

In the case of data on population by locality sizes, we plot cumulated percentage of localities against the cumulative percentage of population in them after arranging the localities in order of size starting with the highest class (either individually or in grouped classes). It seems advantageous not to have too many or too few classes so that the classes have population size large enough for plotting and discrimination.

The cumulative proportion of population X_i is plotted along the X axis and the cumulative proportion of localities Y_i along the Y axis. If there be not much concentration of population by locality size, then

Fig 10.TANZANIA LORENZ CURVES 1978 BY ADMINISTRATIVE LEVEL.



the curve will be near a straight line (line of equality) passing through the origin and with an inclination of 45°. The further the curve deviates from this line, the larger will be the concentration. The curve of highest concentration will be depicted by the 2 sides of the triangle lying along the X axis, the Y axis and closed by the 'line of equality'.

Lorenz curve can also be drawn for cumulative percentage of population and cumulative percentage of area (when geographic units are arranged by density).

The least value of concentration is zero when the curve coincides with the 'line of equality' and the largest value of concentration is 1 when the curve forms the triangle with the line of equality and the X and Y axis. In other cases, the concentration will be between 0 and 1.

To measure the concentration ratio, the formula used in Gini concentration ratio = Sum of $X_i\ Y_{i+1}$ - Sum of $X_{i+1}\ Y_i$ where X_i, Y_i have the meaning already indicated. Another measure of concentration is defined as the Duncan index equal to the value of the maximum difference between X_i and Y_i among all the (X,Y) sets. This is also equal to ½ of sum of absolute difference between the uncumulated values of X and Y. This index is equivalent to the index of dissimilarity.

Figure 10 presents the Lorenz curve for Tanzania 1978 by administrative level. The data for the regions is given in Table 12.

Gini concentration coefficient for data on Table 12 is:

$$10000\ GCR = [(4.8)(.3) + (7.5)(2.6) + \ldots + (94.4)(92.1) + (97.4)(100.0)] - [(.2)(7.5) + (.3)(15.8) + \ldots + (84.7)(97.4) + (92.1)(100.0)] = 3928$$

or GCR = .39

It can be checked that it is equal to:

$$[(4.8 - .2) + (7.5 - .3) + \ldots + (97.4 - 92.1) + (100.0 - 100.0)]/10000$$

Table 12. POPULATION BY REGION BY PERCENTAGE OF AREA AND POPULATION ARRANGED BY DENSITY, 1978

Region	Density (Km^2)	% of Pop. x	Cumulated % of pop, X	% of area y	Cumulated % of area, Y
Dar Es Salaam	553	4.8	4.8	.2	.2
Zanzibar	194	2.7	7.5	.1	.3
Mwanza	73	8.3	15.8	2.3	2.6
Kilimanjaro	68	5.1	20.9	1.5	4.1

Table 12 (Contd...)

Region	Density (KM^2)	% of Pop. x	Cumulated % of pop, X	% of area y	Cumulated % of area, Y
Mtwara	46	4.4	25.3	1.8	5.9
Tanga	39	6.0	31.3	3.1	9.0
W. Lake	36	5.7	37.0	3.2	12.2
Mara	33	4.2	41.2	2.4	14.6
Shinyanga	26	7.5	48.7	5.7	20.3
Dodoma	24	5.5	54.2	4.7	25.0
Mbeya	18	6.2	60.4	6.8	31.8
Kigoma	17	3.7	64.1	4.2	36.0
Iringa	16	5.3	69.4	6.4	42.4
Coast	16	2.9	72.3	3.7	46.1
Morogoro	13	5.4	77.7	8.0	54.1
Singida	12	3.5	81.2	5.6	59.7
Arusha	11	5.3	86.5	9.2	68.9
Tabora	11	4.7	91.2	8.6	77.5
Ruvuma	9	3.2	94.4	7.2	84.7
Lindi	8	3.0	97.4	7.4	92.1
Rukwa	7	2.6	100.0	7.9	100.0

6. ELEMENTARY ANALYSIS OF DATA

Derived figures are useful to assist in summarizing and comparing data. In this connection, we have already seen the role of graphic presentation of data. However, because of the subjective nature of pictorial representations and, in any case, we need quantification of the measures, we shall in this section consider a few of the important summary measures useful in demography. These include averages, ratios, rates, percentages and index numbers.

6.1 AVERAGES

From the tabulated data or, better from the graphs, one can usually notice that there are certain values that are frequently present and others that occur less frequently. These values are called measures of central tendency or averages. There are various types of averages - the arithmetic, geometric and harmonic means, the median and the mode. They differ from each other in their properties, method of calculation and their use in summarizing data.

Arithmetic mean

The arithmetic mean, usually known as the mean, is obtained by summing the values of the observations and dividing it by the number of observations. Hence, if $x_1, x_2, x_3, \ldots, x_n$ are the values of n observations, then the arithmetic mean of these n observations is:

$$\bar{x} = (x_1 + x_2 + x_3 + \ldots + x_n) / n$$

Sometimes the calculation can be simplified by considering an origin and scale. Such a dodge reduces the values of the observations to be considered and hence the arithmetic. For example, if a is the origin and c the scale, then the new variables are $y_i = (x_i - a)/c$. First we calculate the mean of the y values by the similar formula as above and then convert it to the mean of x by the formula:

$$\bar{x} = c\bar{y} + a$$

When the x values have associated frequencies, i.e. they occur more than once then the formula for the mean for such is very similar where we multiply each of the x values by their corresponding frequencies and divide the sum of these products by the total frequency. The formula would be:

$$\bar{x} = (f_1x_1 + f_2x_2 + \ldots + f_nx_n)/N \text{ , where } N = f_1 + f_2 + \ldots + f_n.$$

Here also, the selection of an origin and scale could reduce the calculation. With symbols already introduced, the formula would be:

$$\bar{x} = c\bar{y} + a$$

Mean age of a population

The mean age of a population can be obtained by using the above formula either to the single year, five year or other broad age group data. The smaller the interval into which the data is tabulated, the more accurate will the mean be. Since the mean is unaffected by scale transformations of the frequency, we can deal with the percentage distribution than the actual frequency distribution. From table 11 we reproduce the percentage age distribution of females in Zanzibar in Table 13.

Table 13. PERCENTAGE AGE DISTRIBUTION - FEMALES, ZANZIBAR

Age	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60 +
%	19.2	18.4	11.0	10.1	7.6	6.7	5.5	3.8	4.3	2 5	3.2	1.3	6.4
$y = \frac{(x-32.5)}{5}$	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	7

We have chosen the origin somewhere in the middle of the frequency table and the scale is taken as the class interval. Such selection has reduced the values from 2.5, 7.5, ..., 57.5 and the open interval to values like -6, -5, ..., 5 and 7. The last value is actually chosen arbitrarily as it is very difficult to know the exact limites of the last class to determine the mid point and the transformed value. If the frequency is small, then the arbitrariness in the choice of the value will not have much effect on the calculations. The convention is that the open interval should be such that it contains not more than 2 to 3 % of the frequencies.

.../73

The mean of the transformed variable comes out as :

$$\bar{y} = [19.2\ (-6) + 18.4\ (-5) + \ldots + 6.4\ (7)] / 100 = -2.194$$

Hence $\bar{x} = (-2.194) \times 5 + 32.5 = 21.5$

Average household size

The mean household size can be obtained either from the frequency distribution showing the number of one, two, three ... person households in the country or from the aggregate number of persons and households in the country. In the former case, we have to calculate the aggregates and proceed and in the latter we just divide the number of persons by the number of households. Usually we include only private households in such an exercise. In rural Zanzibar Town/West region, there were in 1978, a total of 5940 private households with 22944 persons giving an average of 3.9 persons per private household (See Table 14).

Table 14. HOUSEHOLDS AND POPULATION BY HOUSEHOLD SIZE, Zanzibar Town/ West Region, 1978 - rural

Household size	No. of hh's	Population	Household size	No. of hh's	Population
1	1462	1462	7	331	2317
2	907	1814	8	230	1840
3	812	2436	9	154	1386
4	723	2892	10+	252	2987
5	604	3020	Total	5940	22944
6	465	2790			

Mean number of children ever born or Parity

The mean number of children ever born alive is obtained either from the frequency distribution of women with 0, 1, ... children ever born by the method already described, or from the ratio of the total number of children ever born to the number of women. Usually, this average is calculated by age of the women. For example, the average number of children ever born to women aged 20-24 years is the ratio of

total number of children ever born to women aged 20-24 years to the total number of women in that age group. For Zanzibar west urban, table 15 presents the number of women aged 20-24 years by number of children ever born.

Table 15. NUMBER OF CHILDREN EVER BORN ALIVE

No. of Children	0	1	2	3	4	5	6	7	8 +	Total
No. of women	88	129	158	175	146	78	37	11	10	832

The mean is obtained by multiplying the number of women by their number of children ever born, cumulating these values and then dividing the total by the number of women. Here also we have an open interval at the end of the table and we assumed that the value corresponds to 9. Under this assumption, the total number of children came out to be 2333 which when divided by 832, the total number of women gave the average as 2.8. This average also is called the mean parity and is denoted by the symbol P_{20-24} or simply as P_2 because it pertains to the second age group among the reproductive age span.

The distribution given above in table 15 is known as the parity distribution for women aged 20-24 years.

Mean age of fertility schedule

The mean age of the fertility schedule is obtained from the table giving the age specific fertility rates. Table 16 presents the age specific fertility schedule for Tanzania. Here also we can take an origin and scale, to reduce computations and derive the mean.

Table 16. AGE SPECIFIC FERTILITY RATES, Tanzania 1978

Age, x	Transformed value, $y=(x-32)/5$	No. of women (1)	No. of children born in past year (2)	Specific fertility rate, f_i (3)=(2)/(1)
15-19	-3	877939	118631	.135
20-24	-2	742519	229355	.309
25-29	-1	703549	209295	.297
30-34	0	504793	124200	.246
35-39	1	446638	80466	.180
40-44	2	348706	30855	.088
45-49	3	311951	11515	.037
Total		3936100	804317	1.292

.../75

First we note that the origin is taken at 32 and not 32.5 because the mid point of the age interval 30-34 is 32 in this particular case since the births in the past year was obtained when the women would be half a year younger on the average. From the table, we first get:

$$\bar{y} = [-3(.135) + -2(.309) + \ldots + 2(.088) + 3(.037)] / (.135 + \ldots + .037)$$

$$= -.0853/ 1.292 = -.66$$

Hence $\bar{x} = -.66(5) + 32 = 28.7$

This value is usually denoted by m

Mean age of mother

There is yet another type of mean age one can derive from the same table and that is the mean age of mothers and is obtained by considering the actual numbers of births at each of the ages and not the specific fertility rates.

Here also the choice of the origin and scale reduces the calculations and we get

$$\bar{y} = [-3(118631) + -2(229355) + \ldots + 3(11515)] / (118631 + \ldots + 11515)$$

$$= -847177 / 804317 = -1.05$$

Therefore $\bar{x} = -1.05 \times 5 + 32 = 26.7$

This value is usually denoted by M.

Mean age at marriage

The mean age at marriage may be measured in a number of different ways involving different types of data and different methods and assumptions, with resulting differences in the interpretation of the figures.

Since registration data is grossly inadequate, one usually uses indirect information on marital status to arrive at the mean age at marriage. The most common information obtained in a census is that on civil status from which one can obtain the proportions married at each age.

The singulate mean at marriage is obtained from the proportions reported as single at each of the age from an enumeration and is calculated by first finding the sum of the percentages single through age group 45-49, multiplying the sum by 5 (the age class interval) and adding 1500 which is 100 times the lowest age at which marriage usually takes place in many societies (taken as 15 in this case). If it is younger or older, then the figure to be used will be 100 times the lowest age. Here we have also assumed that by age 49 most of the marriages would have taken place, and those still single would remain so throughout the life span. If, in any particularly country, the highest age is different, then that age could be substituted in the calculations.

.../76

From the sum so obtained, we have to subtract the value obtained by multiplying the average of the percentage single at ages 45-49 and 50-54 by 50 (the mean age of the group). The figure so obtained is then divided by the proportion non single by age 50 which can be derived by subtracting the average percentage still single by age 50 from 100.

From table 17 showing the percentage single by age in Tanzania, we can calculate the singulate mean at marriage as follows:

Table 17. PERCENTAGE SINGLE BY FIVE YEAR AGE GROUPS, FEMALES - Tanzania, 1978

Age	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54
%	62.5	16.1	5.3	2.9	1.9	1.6	1.2	1.6

The singulate mean age at marriage for the given data is:

$$SMAM = \frac{5(62.5 + 16.1 + \ldots + 1.2) + 1500 - (1.2 + 1.6)\ 50/2}{100 - (1.2 + 1.6)/2}$$

$$= \frac{5(915) + 1500 - 70}{98.6} = 19.1$$

Centre of a population

It is interesting to know the average point for the distribution of population within a country or area. The mean point of the population distributed over an area may be defined as the centre of population gravity for the area, i.e. the point upon which the area would balance, if it were a rigid plane without weight and the population distributed thereon, each individual being assumed to have equal weight and to exert an influence on the central point proportional to his distance from the point. The formula for the co-ordinated for the centre of population may be written as:

$$\bar{x} = (p_1 x_1 + p_2 x_2 + \ldots + p_n x_n) / (p_1 + p_2 + \ldots + p_n) \text{ and}$$

$$\bar{y} = (p_1 y_1 + p_2 y_2 + \ldots + p_n y_n) / (p_1 + p_2 + \ldots + p_n)$$

where $\bar{x}$ and $\bar{y}$ respectively are the latitude and longitude of the centre of population and p_1, p_2, p_n are the populations.

Since the surface of the earth is not flat but has a curvature, a slight modification of the formula to account for this curvature is:

$$\bar{x} = (p_1 x_1 + p_2 x_2 + \ldots + p_n x_n) / (p_1 + p_2 + \ldots + p_n) \text{ and}$$

$$\bar{y} = (p_1 y_1 \cos x_1 + p_2 y_2 \cos x_2 + \ldots + p_n y_n \cos x_n) / (p_1 \cos x_1 + p_2 \cos x_2 + \ldots + p_n \cos x_n)$$

The greater the details in which the population figures are available in terms of smaller geographic units, the better the estimates of the co-ordinates of the centre of population would be.

The Median

When there are extreme observations resulting in an extremely skewed distribution and when some of the class intervals are open, another measure of average of a distribution becomes more appropriate and it is the median or the value which divides a distribution into two equal halves. When observations are given and we rank them by size, then the middlemost item is the median (in case of tie or when there are an even number of observations we take the average of the two items to get the median value). For grouped values, the formula for the median can be obtained by simple linear interpolation as:

Median, $MD = L + (N/2-m)c/f$, where

L = lower limit of the median class, i.e, the class containing the 50% frequency; N = total frequency, m = cumulative frequency upto median class; c = class interval, and f = frequency of median class.

The median age of a population data as given in table 13 can be obtained by first cumulating the frequency (in this case the percentages). We note that the class 15-19 contains the cumulated frequency of 58.7, i.e. the 50% frequency lies in that class because the previous class has cumulative frequency of only 48.6. From the formula the median is:

$$Md = 15 + (50-48.6)(5/10.1) = 15.7$$

Similarly the median household size in Zanzibar west as in Table 14 can be obtained as follows: We cumulate the number of households with 1,2,... household members and find the class containing the 50% frequency. In this case it corresponds to 2970 = 5940/2. The cumulative frequency with size 2 is 2369 and with 3 it is 3181. Hence size 3 contains the median.

$$\text{Median} = 3 + (2970-2369)(1/812) = 3.7$$

In a similar way the median parity is from table 15:

$$Md = 3 + (416-375)(1/175) = 3.2$$

Again, the median age of the fertility schedule from table 16 is:

$$Md = 24.5 + (.646-.444)(5/.297) = 27.9 \text{ , denoted as } m'$$

and the median age of women at child birth is:

$$Md = 24.5 + (402158-347986)(5/209295) = 25.8 \text{ , denoted as } M'.$$

In the calculation of the median point for distribution of population, we start with the northern tip of the country and draw a line horizontally. At intervals of 1° each we draw horizontal lines parallel to the first line and note down the population within the enclosed areas. This becomes easy, if figures are available at small areal levels; otherwise we have to take populations at whatever level is available and consider population of areas which are mostly enclosed within two parallel lines. That one which divides the population into two halves between the north and south is the median line. Similarly, we can draw vertical lines starting with the extreme east tip of the country and find the line dividing the population into two halves. The intersection of the two lines gives the median point.

.../78

The mode

This is yet another measure of central tendency and is defined as that value which occurs the most number of times, i.e. the one with the highest frequency. In ungrouped data, it is simply the one with the largest frequency but when the data is given in grouped values, simple linear interpolation would be needed to obtain the modal value.

The geometric and harmonic means

The arithmetic mean, median and mode are frequently thought of as the more important measures of central tendency, because of their wide usefulness, general applicability and ease of computation. Under certain circumstances there are other measures which are more apporriate. For example, in the study of population growth, the geometric mean is more relevant. The geometric mean of n observations $x_1, x_2, \ldots, x_n$ is

$$GM = (x_1\ x_2 \ldots x_n)^{1/n} \text{ or } \log GM = (1/n)(\log x_1 + \log x_2 + \ldots + \log x_n)$$

For data given in frequency tables, the formula becomes

$$\log GM = (1/N)(f_1 \log x_1 + f_2 \log x_2 + \ldots + f_n \log x_n),$$

$$\text{where } N = f_1 + f_2 + \ldots + f_n$$

For calculating the average annual rate of growth of a population, the size of a population at a given point in time etc, the geometric mean method is useful and we shall consider the applications in a later section.

The harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the values and is denoted as:

$$HM = n/(1/x_1 + 1/x_2 + \ldots + 1/x_n)$$

6.2 RATIOS

Demography is a field concerned with occurences of events like marriage, births, deaths, migration, education, employment, household formation and so on. It is possible to measure the incidence of events in many ways. For any demographic measurement of an event we usually want (a) the time reference period (b) the group referred and (c) the type of occurence being measured.

Even though the absolute number of events occuring has great value, many times we need these events to be related either to time, the size of the group or other variables.

A useful summary measure is the ratio expressing the relative size of two numbers. In this numerator-denominator relationship, there are two types - one in which both the numerator and denominator come from the same universe and the other where they are from different universes. Of the former type, we have: sex ratio, age ratio, child woman ratio, dependency ratio and survival ratio. Examples of the latter are: population density and per capita income.

Sex ratio

It is usually defined as the number of males per 100 females. It could be for a population, for a given age group or of events. The sex ratio at birth is defined as the number of male to 100 female births (usually this lies in a narrow range of 101 to 108 and it is noted to be on the lower range in African populations)

From table 4 we calculate the sex ratio of the population in Mainland Tanzania in 1978 as

SR = (8350492)/(8686006) X 100 = 96.1

Similarly the sex ratio of the population aged 20 to 24 years in Zanzibar is from table 6,

SR = (14914)/(18259) X 100 = 81.7

Age ratio

The United Nations defined an age ratio as the ratio of the number of persons of a given sex-age group to the average of the numbers in adjacent age groups (all the groups must be of equal width) multiplied by 100. Another defintion given by US Bureau of census uses in the denominator the average of the three groups, viz the two adjacent and the one being studied. If digit preference is predominant in the data, it has been shown by Ramachandran to be better to define the age ratio with a denominator having a weighted average of the three groups with the adjacent groups having weight unity and the middle group having weight two.

The age ratio at age 20-24 in Zanzibar for males is:

Age Ratio, AR = 2(14914)/(22721 + 14623) X 100 = 79.9 - UN definition

= 3(14914)/(22721 + 14914 + 14623) X 100 = 85.6 US census Bureau

= 4(14914)/(22721 + 2(14914) + 14623) X 100 = 88.8 Ramachandran

Child woman ratio

There are two types of ratios in usual practice - one pertaining to children aged 0 - 4 and the other with children 5 - 9. The first ratio is defined as

CWR = 100(Children aged 0-4 years)/(Women aged 15-44 years)

and the second is defined as

CWR = 100(Children aged 5-9 years)/(Women aged 20-49 years)

The second definition becomes useful especially when age data is poor and mostly it is at the very young ages that biases occur. The ratios are useful in estimating child migration from two census enumerations and for estimating fertility level.

For Zanzibar the child woman ratios are from table 6. Based on children aged 0-4 years:

$$CWR = 100(44687 + 45949)/(24182 + 18259 + \ldots + 10389) = 99.4$$

Based on children aged 5-9 years:

$$CWR = 100(42486 + 43945)/(18259 + 16017 + \ldots + 5990) = 118.4$$

Dependency ratio

This concept is based on the general dependence of children and old people on the working age population for sustenance. It has not much to do with economic dependence, especially in developing countries with children and older persons contributing substantially to the economy.

Age dependency ratio, ADR = 100(Children aged 0-14 + Old persons aged 65)/(Pop aged 15-64 years)

Child dependency is similar to age dependency but with only children in numerator and old age dependency has only old persons in numerator.

For Zanzibar, the age dependency ratio in 1978 from table 6 is:

$$ADR = 100(44687 + 42486 + 28726 + 45949 + 43945 + 26248 + 3118 + 3029 + 1719 + 3165 + 2109 + 3249 + 937 + 3121)/(22721 + 14914 + \ldots + 6529 + 24182 + 18259 + \ldots + 6072) = 113.7$$

Child dependency ratio is:

$$100(44687 + 42486 + 28726 + 45949 + 43945 + 26248)/(22721 + \ldots + 6529 + 24182 + \ldots + 6072) = 104.2 \text{ and}$$

Old age dependency ratio is:

$$100(3118 + \ldots + 3165 + 2109 + \ldots + 3121)/(22721 + \ldots + 6529 + 24182 + \ldots + 6072) = 9.6$$

Survival ratio

There are two types of survival ratios generally in use. They are cohort survival ratio and overall or open ended survival ratios. In the former the ratio of persons of a given closed sex, age group to those found still alive after the lapse of some time is considered whereas in the latter the groups are from open ended age groups, like say those aged 20 years and above etc.

In Mainland Tanzania there were 157272 males enumerated at age 9 in 1967. Assuming a closed population and that age reporting at both points of time are accurate, these persons would be aged 20 years in the next census, i.e. in 1978. This number was reported as 155134. Thus the survival ratio of males aged 9 years for the 11-year period is:

$$SRM_9^{11} = 155134/157272 = .9864.$$

Apparently this is too high and it is suspected that the ratio has been bloated up by the twin factors of preference for digit 0 and avoidance of digit 9 and also perhaps better enumeration in 1978 than in 1967 due to improvements in organization, education etc. Similarly for males aged 9-13 years in 1967 the survival ratio for the period 1967-78 will be the ratio of males aged 20-24 in 1978 to those aged 9-13 in 1967. This ratio comes out as:

571666/(157272+ 16704 + 93001 + 155122 +113131) = .8339

The open ended survival ratios are calculated in similar fashion keeping the numerator and corresponding denominator with population aged x and above. For example those aged 5 and above in 1967 would be aged 16 and above in 1978.

Table 3 presents both types of survival ratios.

Density

Density of population is a simple concept much used in studies relating population size to resources and in ecological studies. Density is usually computed as population per square kilometer or mile of land area rather than of gross area (including land and water). For some purposes, more meaningful densities are obtained for a country or region by relating the size of its population to the amount of agricultural land.

For example in Zanzibar/Pemba with 2460 sq.km and population in 1978 of 476111 the density = 476111/2460 = 193.5 persons per sq.km.

6.3 RATES

Rate is a special type of ratio which measures the likelihood of occurence of a phenomenon within a given universe characterized by the spatial, temporal and perhaps the individual characteristics of the elements of the universe. Sometimes, the term rate is used in an even more restricted sense by limiting the denominator to the population at risk to a particular type of event. Thus rates are used to study the dynamics of change and refer to the occurence of events over a given interval in time and is defined as:

$$\text{Rate of incidence of an event} = \frac{\text{No. of events that occur within a given time interval}}{\text{No. of members of the population who are exposed to the risk of the event during the same time interval}}$$

The concept of person years lived is the ideal way to specify the population exposed to the risk of an event and is simply the product of the number of persons multiplied by the number of years or fraction of years that each person lived in a given place.

Crude Birth Rate

The crude birth rate is defined as the ratio of the total number of births occuring in a given period, usually a year, to the mid period (year) population and is expressed per 1000 population.

Thus CBR = (B/P) 1000, where B is the total number of births in a year and P is the mid year population. According to table 16, there were 804317 births reported in Tanzania during the period 1977 to 1978. The mid year population can be obtained from the 1978 enumerated figure by applying a compound growth formula for the half year before the census. We take a growth rate of 3.2% (the observed annual average growth rate between 1967 and 1978). The mid year population is thus obtained from the formula:

$$\text{Population of 1977-1978} = \text{Population of } 1978(1\ .032)^{-.5}$$

With a total population enumerated of 17512610, the mid year estimated population comes to 17238958 and hence the crude birth rate is:

$$CBR = (804317/17238958)1000 = 46.7$$

The rate so obtained is called crude because it is affected by age, sex and other characteristics of the population. Since births occur only to women, it is better to calculate a rate based only on the female population. It is even better to derive a measure which takes care of the age distribution of women, because the incidence of births is not uniform within the age span.

General fertility rate

Thus we can calculate a birth rate based on women in the reproductive ages, say 15-49 years. This is called the general fertility rate and is defined as:

$$GFR = (\text{Number of births in a year/mid year female population aged } 15\text{-}49)1000$$

The mid year female population can be calculated as the mid year population by using the compound growth formula.

For the data from Tanzania, the number of women aged 15-49 at the 1978 census was 3936100 and the mid year population would be $3936100(1.032)^{-.5}$, assuming the same growth rate for the female population in the reproductive ages as the total population. Using the figures, we obtain the GFR = (804317/3874595)1000 = 207.6
This measure also is not very useful, because even within the reproductive age span, the fertility of women do differ.

Age specific fertility rate

One method of taking account of this fact is to calculate the birth rate at each age. Thus an age specific birth rate is defined as the ratio of the number of births occuring in a year to women in a specific

age to the total number of women in that age. Thus

ASFR at age x = f_x = Number of births occuring in a year to women aged x years/No. of women aged x.

Table 16 has a column giving the age specific rates for Tanzanian women aged 15-19 to 45-49. For example, the age specific fertility rate at age 25-29 is 2099295/703549 = .297.

Total fertility rate

This is a measure of fertility obtained by summing single year of age specific fertility rates or by summing the age specific fertility rates given in, say, five year age groups and multiplying the sum by 5(the age interval). Based on Table 16 the sum of the age specific fertility rates is 1.292 and hence the

$$TFR = 5(1.292) = 6.5$$

Gross reproduction rate

Instead of using all the births irrespective of sex, we use only the female births and calculate the age specific female fertility rates and sum the single year specific rates or sum the five year rates and multiply by 5, then we get another measure of fertility known as the Gross reproduction rate. If however, births are not available by sex of child, but information on sex ratio at birth is known, then the GRR can be calculated as

$$GRR = TFR(\text{Proportion of female to male births})$$

The female proportion among births is the ratio of female to total births. Since no data are available on sex ratio at birth in Tanzania, we can assume a reasonable figure and proceed. Here we took the sex ratio as 102 and hence obtain the female proportion as

$$(100)/(100+102) = .495.$$

Thus GRR = 6.5(.495) = 3.22

The TFR gives a measure of the total number of children a woman entering the reproductive age can be expected to have when she reaches the end of the reproductive span, assuming that current fertility would hold good into the future. Thus this number would be near the completed family size. Similarly, GRR gives the total number of female children a woman entering reproductive age can be expected to have by the end of her reproductive period under the present fertility regimen.

Net reproduction rate

Since not all women entering the reproductive age complete the entire reproductive span and even from birth to start of reproductive age, there is some attrition. a measure which accounts for this depletion of the cohort by the effect of mortality is the net reproduction rate. It is calculated in a similar fashion as the GRR but instead of the weight 5 being used for the 5 year age groups, we instead use a weight depending on level of mortality. The number 5 in the case of 5 year groups and 1 in the case of single year values, indicate that

all women survive from one age to next. Thus if we use the life table functions depicting the attrition from one age to next, instead of the weights of 1 or 5 as the case may be, to obtain a measure of fertility which has taken cognisance of the fact that some women die during the reproductive span 15 to 49 years and also from birth to age 15. In the notation of life tables, then the NRR is defined as:

NRR = sum of product of age specific female fertility rate and life table survival ratio from birth to the specific age

A life table prepared for Tanzania using mortality information from the census gave survival ratios from birth to ages 15-19, ... 45-49. These values combined with female specific fertility rates would produce the NRR as:

NRR = (.135 X 3.58 + .309 X 3.48 + .297 X 3.37 + .246 X 3.24 + .180 X 3.10 + .088 X 2.96 + .037 X 2.80) X .495, where .495 is the proportion of female births, and 3.58, 3.48, ..., 2.80 respectively are the survival proportion of the females aged 15-19, 20-24, ..., 45-49.
The value comes out to be 2.12.

Sex age adjusted birth rate

This measure of fertility is a type of standardized measure accounting not only for the age structure of the population but also for the pattern of fertility observed among groups. The United Nations observed on the basis of fertility schedules of several countries that the relative fertility at ages 15-19, 20-24, ..., 40-44 are in the proportion 1:7:7:6:4:1.

The sex age adjusted birth rate, SAABR is defined as 1000 times the ratio of births occuring in a year to the weighted sum of the female population in the ages 15-44 weighted according to the fertility weights given above. The weights have been so chosen that the SAABR is not much far from CBR.
For the Tanzanian data the SAABR comes out to be:

SAABR = 1000(804317)/(1X377939 + 7 X 742519 + 7 X 703549 + 6 X 504798 + 4 X 446638 + 1 X 348706) = (8044317/16164461) 1000 = 49.8

Being a measure which is independent of age-sex structure, it has great application specially in comparing rates and for projection purposes when specific rates are unavailable.

Crude death rate

The crude death rate is defined exactly similar to the crude birth rate but in the numerator we use the total number of deaths during the time reference period.

Thus, CDR = 1000(Total number of deaths during a year)/(mid year population)

Age specific death rates

Since the incidence of mortality varies from one age to the next, it is advisable to compute rates specific for age. Thus the age specific death rate at age x is defined as:

$ASDR_x$ = 1000(No. of deaths at age x in a given year)/(mid year population at that age)

= M_x.

It is preferable to calculate the rates separately for each sex. At least at very young and old ages, since the intensity of mortality is very high and changes fast, it is advisable to have the rates by single years of age from age 0 to say 4 and from age 60 onwards. For the intermediate ages five year age groups may be sufficient. Sometimes, it may be only necessary to have ten year age groups for the range 10 to 49 years as the rates do not vary very much in this segment of the population. For an age group x to x+ n, the rate is denoted as $_nM_x$ and is defined similarly as above but with x replaced by x to x+n.

Infant death rate

Of special interest to demographers, public health workers and planners in general, is the mortality of infants, since it gives an indication of the level of living of the population, can be controlled with proper planning, affects a sizeable segment of the population and contributes a high proportion of deaths. It is also a waste of human resources.

Infant death rate is defined as 1000 times the ratio of infant (children under age 1, i.e., those aged 0 years) deaths to number of live births in the year.

M_o = IDR = 1000 (No. of deaths to children under 1 year of age in the year)/(No. of live births in that year)

It is advisable to calculate IDR separately by sex.
The rate given above is not a ratio of deaths per 1000 person years lived in infancy during the year; because the denominator is not an estimate of person years, but the number of live births that were registered. It also does not represent the proportion of a years births that died before age one, because the numerator and denominator do not belong to the same universe.

It is thus closer to being a probability than a rate, since the denominator is persons (infants) exposed to deaths beginning at a certain time (birth) rather than the number of person years lived by infants.

Deaths are not evenly distributed through the first year of life; instead a high proportion occurs during the first week and of these a high proportion occurs during the first day. In general, high infant death rate goes with lower proportion of infant deaths at early life period and in high infant death rate countries, infectious and

parasitic diseases contribute a large proportion of the deaths and if we consider a total of 100 infant deaths in a year, then proportionately a large segment would occur after a week or month of birth because it is during the later period of infancy that such exogenous causes play an important role in mortality.

The infant death rate, however, is not a true probability because not all of the infant deaths in a given year occur to births in the same year. Some of these infant deaths belong to births of the previous year.

Separation factor

An adjusted infant death rate, then would be obtained by considering deaths of infants in a given year and dividing it by a weighted sum of births in the two years involved. The weights are known as 'separation factors'. Usually the separation factor for the current years birth will be near .7 and consequently the factor for the last years birth would be: 1 - .7 = .3. Sometimes the factors are taken as (2/3) and (1/3) respectively.

In many enquiries two questions asked are: number of births in past year and the number of these births still alive on census date. To obtain an estimate of infant death rate we can multiply the reported proportion of infant deaths from the above question, i.e., take 1000(number of deaths among last years birth)/(number of births of past year) and multiply the value by 3/2, i.e., the reciprocal of the separation factor for the current years births.

Foetal death and still birth

Foetal death is defined as, 'death(disappearance of life) prior to the complete explulsion or extraction from its mother of a product of conception irrespective of the duration of pregnancy - the death being indicated by the fact that after such separation the foetus does not breathe or show any other evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles'. Still birth is a foetal death occuring after 20 or 28 weeks of pregnancy.

Migration rates

Migration rates like any other rate can be calculated as the ratio of migrant to the population at risk. Usually, the rate is presented per unit of population. Thus in, out and net migration rates are:

I = (Number of inmigrants)/(Population at risk of inmigration)
O = (Number of out migrants)/(Population at risk of out migrating)
N = (Number of net migrants, i.e., the difference between in and out migrants)/(Population at risk)

The real problem is the calculation of the population at risk. We can use the population at the end of the period in most calculations.

Growth rate

A population grows by the addition of births and inmigrants and is depleted by deaths and outmigrants. Thus a growth rate can be defined as:

r = (birth+ inmigration) rate - (death+ outmigration) rate
if migration is nil or if net migration is zero, then the growth rate is:

r = (birth - death) rate . This is the natural growth rate.
If the curve of population growth is linear, i.e., population is growing in arithmetic progression, then the growth rate can be obtained from two population counts by the formula:

$r = (P_n - P_o)/nP_o$, where P_n and P_o denote respectively the populations counted at the end and beginning period and n is the interval between the two counts. This is obtained from the linear, arithmetic or simple interest law of population growth:

$$P_n = P_o\ (1+ rn)$$

If population is growing as per the compound interest or geometric law, then the growth rate is obtained from the formula

$r = \text{antilog}\ (1/n)\log(P_n/P_o) - 1$ and is derived from the equation $P_n = P_o(1+r)^n$

Here logarithm is taken to base 10 or it is the common logarithm. In case log tables are not readily available, then an approximation to the value of r is given by the formula:

$$r = 2(P_n - P_o)/n(P_n+P_o)$$

Since the growth of a population takes place at every instant of time, the geometric or compound interest formula needs adjustment to take account of this fact. The exponential formula given below depicts population growth under this condition:

$P_n = P_o e^{rn}$, where 'e' is the exponential function which is a constant equal to 2.7183. The growth rate under this law of population growth will be:

$r = (1/n)\ln\ (P_n/P_o)$, where ln denotes logarithm to base 'e' or the natural logarithm.

Let us calculate the arithmetic, geometric and exponential growth rate of the population of Zanzibar during the intercensal period 1967 to 1978. The 1967 population was 354400 and that of 1978 was 476111 and the period was exactly 11 years.

Thus the arithmetic growth rate is (476111 - 354400)/11(354400) = .032, i.e., 3.2% per year.

The geometric growth rate is antilog(1/11)log(476111/354400) - 1

= antilog (1/11)log(1.3434) - 1 = antilog(1/11)(.1282) - 1

= antilog(.01166) - 1 = 1.027 - 1 = .027, i.e., 2.7% per year

Calculating the approximation to the geometric growth rate with the given data we obtain the rate as

= 2(4764111 - 354400)/11(476111 + 354400)

= 2(121711)/11(830511) = 243422/9135621 = .027, i.e., 2.7% per year.

We note that the approximation is not very much different from the exact value. However, if the growth rate is very large then these two would differ significantly.

The exponential rate of growth of the population is:

= (1/11) ln(476111/354400) = (1/11) ln(1.34343) = (1/11)(.2952)

= .027, i.e., 2.7% per year.

Even though the exponential growth rate is same as the geometric growth rate up to two decimals, it is always smaller than geometric rate which in turn is smaller than the arithmetic growth rate.

We use the arithmetic law when population is growing by a constant number per year and the geometric law when the growth rate is constant.

Participation rate

There are several measures used in study of characteristics of a population and most of these are in the form of ratios or rates. We shall consider a few of the important ones below.

Enrolment or school participation rate

Even though crude rates may be sometimes useful, it is generally preferable to calculate age, sex and other characteristic specific enrolment or school participation rates. The definition is very straight forward and is the ratio of number of children of a given age sex or other group enrolled in school or participating in education to the total number of children of that group in the population. Sometimes, we get the numerator from the educational system and the denominator from a census or survey and in certain cases we can get both these from an enumeration where a question has been included on this aspect. Instances of including grade are also found.

In the 1978 census of Tanzania the question asked was whether a person was attending or not attending school. From these figures we can obtain age-sex specific attendance rates.

In Zanzibar West region at age 9 there were 1696 male children of whom 1583 were attending school giving an attendance/participation rate of .933 or 93.3%.

Economic activity or labour force participation rate

The proportion of a population of a given age sex or other characteristic group engaged in economic activities or in the labour force gives the economic activity or labour force participation rate. Usually this is calculated by five year of age group, sex, marital status and other relevant characteristics which have effect on participation in economic activities.

In the 1978 census of Tanzania the question on work status asked pertained to what an individual usually did to earn a living. Tabulation was prepared showing persons by work status. For example, in Zanzibar West region out of the 5404 males, it was reported that 5172 were in the work force giving a participation or activity rate of .957 or 95.7%.

Headship rate

The proportion of a given age, sex, marital status or other relevant categories who are heads of households is called the headship rate. It gives an idea of the characteristic of the household and is useful in estimation of housing needs.

In the 1978 census of Tanzania, the question on relationship of the members of a household and the characteristics of the head were collected, as the questionnaire was for the private households, one which took the household as the unit of enumeration.

In Zanzibar West region there were 593 in the rural and 982 in the urban areas out of a total 5404 males who were reported as heads of households at age 25-29 giving a headship rate of 25.1 at the age for males.

Literacy rate

The proportion of a given age, sex or other relevant categories who are considered able to read and write in a specified language are called literates.

In Tanzania in the 1978 census the question referred to ability to read and write Kiswahili. Tabulations were prepared giving numbers of persons by age and sex who were literate and illiterate for those aged 5 years and above. In Zanzibar West region there were 4566 literates among males aged 25-29 years out of a total 5404 males giving a literacy rate of .845 or 84.5%.

Consumption rate and adult equivalents

There are certain characteristics like utilization of health/hospital facilities, consumption of food or other consumer goods and services which are used by the population but with varying intensity. For example, it is known that the utilization or need for hospital or

health services differ from one sex to another, from one age to the next and between socio-economic and geographic groups. For projection purposes, we need these types of rates or units.

For example, it may be possible to have the rate of utilization or need for health facilities by age, sex and other characteristics. Similarly for food consumption or needs we can have the rates by age, sex and other relevant categories.

Adult equivalents are obtained from consumption/utilization rates by taking the figure for the adult as unity and obtaining the relative values for the others.

6.4 PROPORTION

Proportion is a special type of ratio in which the denominator includes the numerator. We have already seen some of these proportions in connection with definition of literacy. We shall now consider a few more which are important in demographic analysis.

Sex proportion

The sex proportion is usually defined as the ratio of males per 100 total population. Thus the sex proportion is the ratio of number of males to total number of males and females and is usually multiplied by 100. This porportion can be calculated by age, geographic or other characteristics.

The sex proportion of the population in Tanzania, 1978 is

$SP = (8350492)/(8350492 + 8686006) = .49.$

This can also be obtained from the relationship

$SP = SR/(100 + SR) = 96.1/196.1 = .49$

The sex proportion gives an indication of the composition of the population by sex.

Urban proportion

The urban proportion is defined as the ratio of the urban to the total population. This may be calculated by age, sex and residence. The urban population in Zanzibar in 1978 was 154980 out of a total of 476111 so that the urban proportion is:

$UP = 154980/476111 = .33$, i.e., 33% of the population of Zanzibar in 1978 was defined as urban.

6.5 PERCENTAGE

Percentage is yet another special type of proportion in which the ratio is multiplied by 100 so that the ratio is expressed per 100. Ratios and percentages are useful for analysing the composition of a set of events or of a population.

We have already been familiar with percentages while considering some of the proportions like urban proportion etc.

One of the most important application of percentage distribution is that of the age distribution. It is usually carried out by sex, geographic characteristics and by age or age groups. In calculating the percentages, it is advisable to exclude the unknown or age not stated category, provided that this category is not very large, say, less than 1%. Table 11 gives the percentage age distribution for Tanzania. The calculation of the percentages at any age, sex group is straight forward. For example, the percentage of males aged 25-29 is 100(number of males aged 25-29)/Number of males in the country)

$$= 100(610325)/(8587036) = 7.71$$

Allied to the percentage age distribution is the cumulated percentage distribution which is obtained by cumulating the percentages from the top downwards. These figures in the table on page 1?y are obtained from table 11. These are of great value in the calculation of the percentages including the median and in the application of the Coale-Demeny method in the estimation of vital parameters from age distribution.

The percentage distribution of the fertility schedule and the cumulated form are important in the study of the age pattern of fertility.

6.6 PROBABILITY

Rates and ratios provide us with a set of useful measures for answering questions on mortality, fertility and other aspects of a population. In the study of mortality especially there are several questions needing answers which require the introduction of other measures. One of the most important measure in mortality is the expectation of life at birth and other indices of mortality which come from life tables. In order to introduce the concepts and terms in a life table, we need to introduce the concept of probability. A probability is similar to the rate with the denominator composed of all those persons in the given population at the beginning of the period of observation. Since it may be difficult to have such data, we try to utilize measures already known to arrive at the require probability.

One of the most important element in a life table is the probability of survival from an exact age for a period of one or five years. This is denoted by p_x or ${}_5p_x$ respectively. The complement, i.e. the probability of dying is obtained by subtracting the value from unity. Hence q_x or ${}_5q_x$ is equal to respectively $1-p_x$ or $1-{}_5p_x$. Approximately q_x or ${}_5q_x$ can be obtained from the age specific death rates M_x or ${}_5M_x$.

The relationship is $q_x = 2M_x/(2+M_x)$

$${}_5q_x = 2(5)\ {}_5M_x/(2+5{}_5M_x) \text{ or generally,}$$

$${}_nq_x = 2n\,{}_nM_x/(2+n\,{}_nM_x)$$

From these probabilities we can obtain the other life table functions as follows: if l_0 is the radix, i.e. the hypothetical number of births then the survivors of these births can be obtained by multiplying the number of births by the corresponding probability of survival to obtain the numbers surviving at various ages. For example

$$l_1 = l_0 p_0$$

$$l_5 = l_0({}_5p_0)\ ,\ l_{10} = l_5({}_5p_5)\ \ldots\ldots$$

Thus l_x is the number of survivors at exact age x.

The number of deaths can then be obtained by differencing the values of l_x at consecutive ages. For example $d_0 = l_0 - l_1 = l_0\,q_0$; $d_x = l_x - l_{x+1}$. Generally, ${}_nd_x = l_x - l_{x+n}$.

Another important concept in a life table is the years lived between exact age x and x+n defined by the symbol ${}_nL_x$ or $L_{x,\ x+n}$. For ages 5 to 75, this can approximately be obtained from the formula:

$${}_nL_x = (n/2)(l_x + l_{x+n}).$$

For the very young ages since the curve of mortality changes drastically, a linear assumption, as implicitly implied in the formula already quoted for ages 5 to 75 will not be appropriate. Hence some modification will be needed to account for this fact and we obtain:

$$L_0 = .33\ l_0 + .67\ l_1$$

$${}_4L_1 = 1.4\ l_1 + 2.6\ l_5$$

The problem at the older ages is similar. The curve of mortality falls steeply and a linear trend cannot be justified. Since it is not necessary usually to have the life table values beyond age 75 or 80, we terminate the life table say at age 80 and consider only the group 80 years and above.

The formula for the years lived after age 80 is denoted by

${}_{\infty}L_{80}$ or $L_{80}+$ or T_{80}, and the formula for $L_{80}+$ or T_{80} is

$l_{80}\ (3.725 + .0000625\ l_{80})$.

We have already seen the symbol T_{80} which is equal to $L_{80}+$ or the total years lived after age 80. We can define T_x as the total years lived after age x which can be seen to be equal to the sum of L_x values from age x onwards. A special value of T_x is the value at age 0, i.e., T_0 and this gives the total years lived at age zero.

An important function derived from a life table is the average number of years lived by a hypothetical cohort of births which is implied by a given life table. This average can be obtained as the ratio of T_0 to l_0 value. At other ages the average is obtained by dividing the T_x value by the l_x value. The ratio is denoted by the symbol e_x. Thus the value of e_0 which is the average number of years lived by a

hypothetical birth cohort obtained by dividing T_0 by l_0 is known as the expectation of life at birth or at age 0. The values of e_x at other values of x denote the average number of years lived by a person aged exactly x years.

In addition to measuring mortality and providing an index of level of living as shown by the expectation of life at birth, a life table enables one to answer queries on the survival probabilities of persons at a given age for the next n years, the proportion of a given age group surviving upto a given age, for comparing mortality experiences of countries at a given time or of a country over time period, it also provides tool for several analytical methods like studying widowhood, orphanhood, length of working age, marital life etc. Life tables also are basic in the construction of stable population models, for population projections and estimation purposes.

6.7 STANDARDIZATION

A crude rate is not a reliable index and more detailed measures like age specific rates are so detailed as to make them cumbersome for interpretation. Occasionally, brevity is essential and a single accurate figure is needed as a basis for comparison. This is secured by considering the detailed information for a population to a standardized or adjusted rate. This manipulation has some advantages of both crude and specific rates while avoiding some of the disadvantages. It summarizes a set of age specific rates independently of age composition and at the same time shows the probable influence of a population age composition on its crude rates when its actual age specific rates are not known. These are achieved by the two methods - the direct and indirect method of standardization. In the method of direct standardization, we apply age specific rates to a standard population and in indirect standardization we use rates from a standard population with age distribution of the given population.

Thus standardization serves the purpose of refining comparison by statistically eliminating an extraneous source of variation in the data that may seriously affect the analysis of the subject under investigation.

We have already seen the sex age adjusted birth rate and the total fertility rate (the gross reproduction rate being related to TFR through the sex proportion at birth is also included) as examples of some sort of standardization. We shall take up some illustrations of direct and indirect standardization based on data from Tanzania.

The crude birth rates in Zanzibar and Mainland based on reported births of the past year came to 48.0 and 45.9 respectively. Thus Mainland apparently has a birth rate 5% lower than that of Zanzibar. In order to see the effect of age-sex distribution on the rates, we can standardize the rate by direct method. For this purpose we use the age specific fertility rates of Mainland and the age distribution of females in Zanzibar as given in Table 6. With these 2 sets of values, we can calculate the expected number of births in Mainland had it

possessed an age-sex distribution like that of Zanzibar. The expected number of births is the product of age specific fertility rates and the corresponding number of females, i.e.

$$\text{Expected No. of births} = (.132)(24182) + (.308)(18259) + \ldots$$

$$+ (.037)(5990) = 19694$$

$$\text{Standardized birth rate} = \frac{\text{Expected births}}{\text{Total population of Standard Population}}$$

$$= \frac{19614}{476111} = 41.2$$

This is 11% lower than the crude birth rate in Mainland. It shows that if Mainland had a similar age-sex composition as Zanzibar, the crude birth rate would have been even lower than the 5% difference noted by the crude rates.

Thus for the application of the direct method of standardization we need age specific rates for the population for which we are calculating the standardized rate and also we need an age distribution of a standard population. It could be the age distribution of the population with which we are trying to compare the rates or it could be a third population.

When we do not have specific rates as in the case of mortality in Tanzania, but have only a crude rate, then we have to utilize the indirect method of standardization. In this method, we use the age distribution of the population for which standardized rate is being calculated along with specific rates for a standard population. The method is to calculate the standardized rate by multiplying the specific rates of the standard population by the age distribution of the population for which the standardized rate is being calculated.

The crude death rate in Mainland Tanzania has been estimated as 19.1. This is for both sexes. Since male mortality has been observed to be higher than female, the crude death rate for male may be higher than 19.1. Assuming that the age specific male death rates of Sweden in 1980 is used as a standard along with the age distribution of male population in Tanzania as from the 1978 census, we can obtain an expected number of deaths and then an indirect standardized death rate which can then be compared with the reported crude death rate to see the effect of age distribution on the death rate. Table 18 gives the age specific death rates for males in Sweden as of 1980.

Table 18 AGE SPECIFIC DEATH RATES, MALES - SWEDEN, 1980 (per 000)

Age	0	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44
ASDR	8.1	0.4	0.3	0.2	0.6	1.0	1.2	1.3	1.8	2.7
Age	**45-49**	**50-54**	**55-59**	**60-64**	**65-59**	**70-74**	**75-79**	**80-84**	**85+**	**All ages**
ASDR	4.4	6.8	10.7	17.9	2[illegible].4	47.0	75.9	123.7	222.0	12.1

Using the age distribution of males in Tanzania Mainland in 1978 as given in Tables 4,6 we get the expected number of deaths as.

Expected number of deaths = (8.1)(294731) + (0.4)(1218561) + ... + (123.7)(35498) + (222.0)(50119)/1000 = 44021

Reported deaths = (19.1/1000)/(8350492) = 159494

$$\text{Age adjusted death rate} = \frac{\text{Reported deaths}}{\text{Expected deaths}} \times \text{Death rate of standard population}$$

$$= \frac{(159494)}{44021}(12.1) = 43.8$$

Thus, instead of the 58% difference shown by the crude death rates of Mainland and Sweden, the standardized rate indicates a 262% difference. The lower difference is due to the favourable age distribution in Mainland.

6.8 INDEX NUMBERS

There are various summary measures which are used in the evaluation, analysis and interpretation of demographic data. Some of these indices will be considered here.

Indices of digit preference

As mentioned in section 4.4, one of the major errors in single year of age data is that due to digit preference in reporting. Even though the incidence of this error or bias can be noted either from the data itself or more vividly from the single year age pyramid or histogram, it is still necessary to arrive at an index for measuring this phenomenon so that it could be used for comparative purposes by sex, over time, geographic area, socio-economic characteristics, etc.

(i) Whipples Index

This index is one of the simplest summary measure of incidence of preference for digits. Typically, it is calculated to measure the strength of preference for digit 0 or 5 or both since, in most cases, it has been noted that 0 is by far the highest preferred digit in age reporting followed by preference for digit 5. Since other types of

errors like age preference, age shifting, selective under or over enumeration are found in young and or old ages we consider only ages 23 - 62 in the calculation of Whipple's index. It is calculated separately for each sex.

For measuring the strength of preference for digit zero, the index is defined

$$\frac{P_{30} + P_{40} + P_{50} + P_{60}}{\frac{1}{10}\left(P_{23} + P_{24} + P_{25} + \cdots P_{60} + P_{61} + P_{62}\right)} \times 100$$

Where P_x = Population of a given sex aged X years.

A similar index can be defined for measuring the preference for digit 5. Here a value of 100 indicates no preference and 1000 indicates total preference for the digit. For measuring the incidence of preference for digits 0 and 5, the index is defined as:

$$\frac{P_{25} + P_{30} + P_{35} + \cdots P_{55} + P_{60}}{\frac{1}{5}\left(P_{23} + P_{24} + \cdots P_{61} + P_{62}\right)} \times 100$$

This index will lie between 100 (no preference) and 500(total preference).

From Table 4 we can calculate Whipples index for rural males in the Mainland as:

$P_{25} + P_{30} + P_{35} + P_{40} + P_{45} + P_{50} + P_{55} + P_{60}$ = (116752 + 129552 + 105580 + 101759 + 98302 + 84514 + 53826 + 76182) = 766467

The population aged 23 to 62 can be written as:

$P_{23} + P_{24} + P_{25-29} + P_{30-34} + P_{35-39} + P_{40-44} + P_{45-49} + P_{50-54} + P_{55-59}$

$P_{60} + P_{61} + P_{62}$ = 67552 + 85261 + 459213 + 354808 + 353208 + 263108 + 271465 + 181256 + 76182 + 13910 + 20082 = 2347700

Hence Whipples index, 100(766467)/(2347700/5) = 163.2

Table 5 gives the values for females and urban areas. From the table we note that the index is very large - more for females and for rural population.

As can be noted, the Whipples index is generally calculated for measuring preference for digits 0 and 5 which are usually attractive digits for reporting ages. But, at the same time, there is need for combining in one index, a measure of the attraction of preferred digits. This is to a certain extent taken care by the following index which is called as 'Digit Preference Index' (DPI).

(ii) The Digit Preference Index (DPI)

In this index, first we calculate the percentage reported in each of the terminal digits 0, 1, 2, ... 9 in the age range 10-69. (This age range is used to avoid the types of other errors indicated as in the case of Whipple's index). If there had been no preference for any specific digit, then each of the percentage could be near 10. Deviation from 10 indicates departure from no preference and the index of digit preference is defined as the sum of absolute deviations of these percentages calculated at each of the digits.

For example, for digit '2' the percentage is calculated as

$$\frac{P_{12} + P_{22} + P_{32} + \cdots + P_{52} + P_{62}}{P_{10} + P_{11} + \cdots + P_{69}} \times 100 = D_2$$

Similarly we can calculate D_o, D_1, ... D_9.

The DPI = Sum of $/D_o - 10/ + /D_1 - 10/ + \ldots + /D_9 - 10/$

The value of the index will be between 0 (no preference) and 180 (one specific digit is preferred over all others).

(iii) Myer's Index

In the calculation of DPI the starting digit 0 gets an undue advantage. Myer's index is similar to DPI but gives equal chance for all digits to get all positions. In other words, in Myer's index ten matrices are considered instead of one matrix starting with P_{10}.

As in DPI the numbers reported in 0, 1, 2, ...9 are calculated first in the age range 10-59. Then similar calculation is done for age range 11-60, 12-61, 13-62, ... 19-68. In other words all digits 0, 1, 2 ...,9 get all possible positions in the arrangements and hence do not have undue advantage/disadvantage. Then percentages are calculated at end digits 0, 1, 2, ...9 and deviation from 10 summed up irrespective of sign to obtain the Myer's index. Here also the value will be between 0 (no preference) and 180 (total preference for one digit).

The calculation can be simplified if one uses the following scheme:

Age range	Population with digit									
	0	1	2	3	4	5	6	7	8	9
10-59	A_o	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9
20-69	B_o	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9

..., 98

For example $A_3 = P_{13} + P_{23} + P_{33} + P_{43} + P_{53}$

$B_6 = P_{26} + P_{36} + P_{46} + P_{56} + P_{66}$ and so on

From the set of (A, B) values we calculate blended sums as follows:

Digit	First Sum (1)	Weight (2)	Second Sum (3)	Weight (4)	Blended Sum (1) x (2) + (3) x (4)
0	A_0	1	B_0	9	C_0
1	A_1	2	B_1	8	C_1
2	A_2	3	B_2	7	C_2
3	A_3	4	B_3	6	C_3
4	A_4	5	B_4	5	C_4
5	A_5	6	B_5	4	C_5
6	A_6	7	B_6	3	C_6
7	A_7	8	B_7	2	C_7
8	A_8	9	B_8	1	C_8
9	A_9	10	B_9	0	C_9
Sum					C

For example $C_5 = 6\ A_5 + 4\ B_5$ etc.

C = Sum of C's $= C_0 + C_1 + C_2 + \ldots\ C_9$

Blended percentage at digit 'i' $= \frac{C_i}{C} \times 100 = D_i$

Myer's index = Sum of absolute deviations $D_i - 10$

$= /D_0 - 10/ + /D_1 - 10/ + \ldots + /D_9 - 10/$

From Table 4 we can calculate the values of A and B. We illustrate the calculation of A_1 and B_1 i.e., the sums of populations with end digit 1 in the age ranges 10-59 and 20-69 for males in rural Mainland Tanzania as:

$A_1 = P_{11} + P_{21} + P_{31} + P_{41} + P_{51}$ = 146917 + 74495 + 54237 + 32393 + 23030 = 331072

$B_1 = P_{21} + P_{31} + P_{41} + P_{51} + P_{61}$ = 74495 + 54237 + 32393 + 23030 + 13910 = 198065

Similarly we can calculate all the values of A and B as in the following table:

Digit	A	Weight	B	Weight	Blended Sum	%	Deviation from 10
0	653882	1	516824	9	5310298	15.2	5.2
1	331072	2	198065	8	2246664	6.5	-3.6
2	476394	3	280486	7	3392594	9.7	-0.3
3	354636	4	197317	6	2602446	7.5	-2.5
4	369790	5	220806	5	2952980	8.5	-1.5
5	540426	6	416803	4	4909768	14.1	4.1
6	379802	7	248203	3	3403223	9.8	-0.2
7	299045	8	188797	2	2769954	7.9	-2.1
8	469245	9	329370	1	4552575	13.1	3.1
9	271746	10	194739	0	2717460	7.8	-2.2
Sum			34857952				

Results for females and for urban areas are given in Table 5.

United Nations Age accuracy Index (Joint Score)

When age data are presented in five year or ten year groups, an idea of the quality of the data can be observed by calculating age ratios and sex ratios. Fluctuations of age ratios from 100 and of sex ratios from one age to another will give indications of age reporting and other errors.

United Nations defined an age ratio score as the average of the sum of absolute deviations of age ratios from 100 and a sex ratio score as the average of the absolute deviations of consecutive sex ratios. Thus we will have 3 scores:- Male age ratio, Female age ratio and sex ratio scores. (MARS, FARS & SRS). United Nations derived a combined measure of age accuracy based on these 3 scores and defined it as: U.N. Joint = Age accuracy Index

$$= \text{MARS} + \text{FARS} + 3\ \text{SRS}$$

A value of UN Joint Score more than 40 indicates poor quality of data. A value between 0 - 20 is considered acceptable and between 21 - 40 as usable with some adjustments. In cases where it is very high i.e. more than 60 or so, as in the case of the data for the country in 1967 or for Zanzibar even in 1978, great care and caution is called forth

in the manipulation of the age sex related data and every effort should be directed to improve age - sex reporting in future data collection efforts.

Replacement Index

This measure related to reproductivity in the five years preceding the observed age distribution and hence is defined as the ratio of child woman ratio in the observed population to the corresponding ratio in the life table or stationary population. (Basing on children under age 5). Similar measure with other groups can also be defined.

$$\text{Replacement Index } J = \left(\frac{P_{0-4}}{W_{15-49}}\right) \div \left(\frac{L^m_{0-4} + L^f_{0-4}}{L^f_{15-49}}\right)$$

Where P_{0-4} = Children aged 0-4 (both sexes)

W_{15-49} = Female population in reproductive age 15-49

L^m_{0-4} = Male life table population aged 0-4

L^f_{0-4} = Female life table population aged 0-4

L^f_{15-49} = Female life table population aged 15-49

Index of dissimilarity

A summary measure of the difference between two age distributions is based on the absolute differences between the percents at each age. The index is obtained as the sum of positive differences or half the sum of absolute differences. We have already referred to the Duncan index and mentioned it as equivalent to the index of dissimilarity. This is yet another instance of the application of the index. It can also be used in comparing two percent distributions, say of areas by race etc.

7. SIMPLE INTERPOLATION WITH EQUAL AND UNEQUAL INTERVALS, EXTRAPOLATION

DEFINITION:

Interpolation is defined, in a narrow sense, as the art of inferring intermediate values in a given series of data by use of a mathematical formula or a graphic procedure.

Extrapolation is the art of inferring values that go beyond the given series of data by use of a mathematical formula or a graphic procedure.

Many of the techniques used for interpolation are suitable for extrapolation. Hence the term interpolation is commonly used to refer to both interpolation and extrapolation. Interpolation is not only used

for estimating intermediate values (e.g. annual population figures from decennial figures) but also for subdividing grouped data into component parts (e.g. figures for single years of age from data for five-year age groups) and for inferring rates for sub-groups from rates of broad groups.

NATURE AND TYPE OF INTERPOLATION

There are various methods of interpolation. It is a matter of individual judgement whether the method selected is suitable and whether or not the approximation is of useful quality. For extrapolation, there is even less basis for judgement in the degree of error. Extrapolation is generally more risky than interpolation. The results are subject to large errors because past trends are not always indicative of what will happen in the future or the relationships in one part of the range of the data may not be a good indication of relationships in another part.

Interpolation by mathematical formula has the quality of imputing a regularity or smoothness to the given series of data or even imposing these characteristics on the data. The regularity imputed or imposed may be unreal. True fluctuations in data on population may exist due to past variations in births, deaths, and migration, especially if there have been wars, epidemics, population transfers, refugee movements etc. Interpolation may usefully serve to adjust defective data, even though some real fluctuations are removed, or to eliminate abnormalities from a series, such as those due to war, when the underlying pattern or trends is wanted.

7.1 INTERPOLATION METHODS

Graphic Interpolation

For many purposes it may be enough to plot a series of given data on a large-scale graph, draw a free-hand curve through the plotted points, and then read off the intermediate points from the graph as needed. However, this method does not satisfy certain requirements particularly for those people working with large amounts of demographic data. For one thing, the requirement of reproducibility is not met. Different workers will produce different results with the hand-and-eye technique, even though such differences may not be large. For another, the computability requirement is not met. The access to desk calculators and electronic computers has generated a demand for computable interpolation procedures which can take advantage of the speed and economy of these machines. Another disadvantage of graphic interpolation is that it may be difficult to read off more than two or three significant figures from the graph. Data that depict a linear trend are easier to plot than those having a non-linear trend and therefore it has been the practice to transform the latter type of data into a form that yields a nearly linear trend. The methods of data transformation include the use of logarithmic scale, plotting of reciprocals or the square root of the original data. There are statistical tools, which are fairly simple, that can help to detect the type of information that will yield a linear or a close to near trend.

.../102

Polynomial Interpolation

Polynomial interpolation is interpolation where the series is assumed to conform to an equation of the general type:

$$y = A + Bx + Cx^2 + Ex^3 \ldots$$

More or fewer terms may be used. A straight line or linear equation will be of the form

$$y = A + Bx$$

and since there are 2 constants A and B in the equation, this can be obtained with two points. A quadratic, or parabola, which requires three points for it to be obtained is of the form

$$y = A + Bx + Cx^2$$

More generally a polynomial equation of the nth degree requires n+1 points. The choice of the degree of polynomial would depend on the nature of the data to be interpolated. A method of polynomial interpolation is discussed below.

Waring's Formula

This is also known as the Lagrange Formula or the Waring-Lagrange formula. The Waring formula is used to derive the multipliers to interpolate for the f(x) value corresponding to a given value of x

The formula for interpolation between four points by a polynomial is

$$f(x) = f(a)\frac{(x-b)(x-c)(x-d)}{(a-b)(a-c)(a-d)} + f(b)\frac{(x-a)(x-c)(x-d)}{(b-a)(b-c)(b-d)} + f(c)\frac{(x-a)(x-b)(x-d)}{(c-a)(c-b)(c-d)} + f(d)\frac{(x-a)(x-b)(x-c)}{(d-a)(d-b)(d-c)}$$, where

f(a) is the value of the function when x = a i.e.

f(a) is the observed value of the y ordinate for the abscissa x = a. Similarly for x = b, c and d.

This is equivalen to the polynomial

$$y = A + Bx + Cx^2 + Ex^3 = f(x),$$

passing through the four points f(a), f(b), f(c) and f(d) to derive f(x). The points do not have to be equally spaced. Interpolation along a straight line would use the formula

$$f(x) = f(a)\frac{(x-b)}{(a-b)} + f(b)\frac{(x-a)}{(b-a)}$$

This formula enables interpolating a value f(x) which lies between f(a) and f(b), i.e. for values of x between a and b. For example the population of Zanzibar in 1957 was 299,111 while that for 1967 was 354,815.

Suppose we want to estimate the population of Zanzibar in 1960. Then using the Waring's formula for two points we have

$$\text{Population in 1960} = f(1960) = f(1957)\,\frac{(1960-1967)}{(1957-1967)} + f(1967)\,\frac{(1960-1957)}{(1967-1957)}$$

$$= 299{,}111\,\frac{(-7)}{(-10)} + 354815\,\frac{(3)}{(10)}$$

$$= 299{,}111\ (0.7) + 354815\ (0.3)$$

$$= 315822$$

Suppose we apply this to the 1957 and 1978 figures to estimate the 1967 figure. The Zanzibar population for 1978 was 476,101

$$f(1967) = f(1957)\,\frac{(1967-1978)}{(1957-1978)} + f(1978)\,\frac{(1967-1957)}{(1978-1957)}$$

$$= 299{,}111\,\frac{(-11)}{(-21)} + 476{,}101\,\frac{(10)}{(21)}$$

$$= (299{,}111)(0.5238) + 476{,}101\ (0.4762)$$

$$= 383{,}391$$

i.e. by using Waring's two point formula the population of Zanzibar was 383,392. Compare this with the actual census figure of 354,815. Using census figures for 1957, 1967 and 1978 a three point formula can be used for estimating population figures for the years between 1957 and 1978.

From table on page 120a, calculate the multiplying factor corresponding to $f_1/f_2 = .429$ for age 15-20. The table shows the following multiplying factors:

$f_1/f_2 = .330$ with multiplying factor 1.950

$f_1/f_2 = .460$ with multiplying factor 2.305.

Hence multiplying factor for $f_1/f_2 = .429$ will be

$$1.950\frac{(.429-.460)}{.330-.460} + 2.305\frac{(.429-.330)}{.460-.330}$$

7.2 GRADUATION AND SMOOTHING

Graduation or smoothing is another type of interpolation designed to obtain a smooth series of values from an irregular series of observed values. The smooth series thus obtained is taken to represent some underlying law governing the behaviour of the observed data. Graduation is commonly used by demographers in connection with the adjustment of series on the age distribution of population.

One of the simplest way of graduating or smoothing of a set of data is by graphical means. But, this method is subjective and not much accuracy can be expected. Algebraic methods like fitting curves, interpolation formulas etc. are also available for this purpose. We shall now consider one group of formulas based on the concept of moving averages.

In the analysis of time series data the method of moving averages is usually applied to remove fluctuations and come out with a set of smoothed values. Even though in demographic analysis the conditions for the application of the method may not be fully satisfied, (in addition to random fluctuations there could be high systematic errors), still the methods seem to work tolerably well when the magnitude of errors of a systematic nature (like digit preference error) are not large. Again if one knows the age range within which the error may be occuring, one could consider these ages for calculation of the moving average (e.g. a ten year cycle may remove some of the digit preference errors as all digits are included in the calculation of an average). When five year of age data is considered, then perhaps the inclusion of one or two adjacent age groups might be sufficient to account for shifting etc.

The U.N. 3 point formula

This is obtained by iterating a moving average straight line twice. For example, if x_1, x_2, x_3, x_4 .... are equidistant values, then the moving average obtained by fitting a straight line through 2 consecutive points will give a smoothed value for a midpoint. Thus

$\bar{x}_{1.5}$ (the smoothed value between x_1 and x_2) $= \dfrac{x_1 + x_2}{2}$

similarly $\bar{x}_{2.5} = \dfrac{x_2 + x_3}{2}$ and so on.

Now if we fit a moving average straight line to these smoothed values, then

$$\bar{x}_2 = \frac{\bar{x}_{1.5} + \bar{x}_{2.5}}{2} = \frac{x_1 + x_2}{4} + \frac{x_2 + x_3}{4}$$

$$= \frac{x_1 + 2x_2 + x_3}{4}$$

The formula thus is written as $\frac{1}{4}$ (1,2,1) where the weights are $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ for the 3 consecutive points and the mid value is smoothed.

It is very suitable to apply this formula to smooth data given in five year age groups, especially when migration or other factors also might have affected the age distribution in addition to digit preference, age shifting etc.

The U.N. five point formula

In this method five consecutive equidistant values are utilized to arrive at the smoothed mid value. This is also a moving average method and is based on similar premises as the 3 point version given earlier. The formula is written as

$\frac{1}{16}$ (- 1, 4, 10, 4, - 1) where the weights are

$\frac{-1}{16}$, $\frac{4}{16}$, $\frac{10}{16}$, $\frac{4}{16}$, $\frac{1}{16}$ for the five consecutive values to arrive at the smoothed mid value.

.../105

This formula also is useful for smoothing five year of age group data but since it uses five points it is better if migration or other factors have not affected the age distribution.

A milder version of the UN Secretariat formula given above is based on the formula: $\frac{1}{36}\left\{-1, 4, 30, 4, -1\right\}$.
As can be seen, this formula gives more weight to the value at the middle and hence is useful when error are not very drastic.

From Table 6 showing the five years of age group data for males in Tanzania we can obtain smoothed values using the above formulae as follows:

Three point formula

$$\text{Smoothed population aged 25-29 years} = \tfrac{1}{4}\left[P_{20-24} + 2\,P_{25-29} + P_{30-34}\right]$$

$$= \tfrac{1}{4}\left[586580 + 2(610325) + 457537\right] = 566192$$

U.N. Five point formula

Smoothed population aged 25-29 years

$$= \frac{1}{16}\left[-P_{15-19} + 4\,P_{20-24} + 10\,P_{25-29} + 4\,P_{30-34} - P_{35-39}\right]$$

$$= \frac{1}{16}\left[-841340 + 4(586580) + 10(610325) + 4(457537) - 439515\right]$$

$$= 562429$$

Five point formula (Mild version)

Smoothed population aged 25-29 years

$$= \frac{1}{36}\left[-P_{15-19} + 4\,P_{20-24} + 30\,P_{25-29} + 4\,P_{30-34} - P_{35-39}\right]$$

$$= \frac{1}{36}\left[-841340 + 4(586580) + 30(610325) + 4(457537) - 439515\right]$$

$$= 589038$$

Newton's halfing formula

This is actually an interpolation method to split data given in 10 year age group into five year groups. As a smoothing technique it is applied when single year or five year of age data are found to be affected by digit preference, age preference, age shifting and other types of age reporting/recording errors.

Single year or five year data are combined to form ten year groups and these are then split into five year groups by the Newton's halfing formula:

for the first half: $\frac{1}{16}\begin{pmatrix} 1, 8, -1 \\ \end{pmatrix}$

for the second half: $\frac{1}{16}$ (- 1, 8, 1)

Thus, if we have 3 ten year age groups, then the middle group can be split into five year groups by using weights $\frac{1}{16}$, $\frac{1}{2}$ and $-\frac{1}{16}$ and $-\frac{1}{16}$, $\frac{1}{2}$ and $\frac{1}{16}$ respectively to get the first and second five years groups.

Carrier - Farrag ratio method

Another method to smooth age data by grouping first into ten year groups is given by Carrier and Farrag. Schematically the method is illustrated below.

Age Group (1)	Five year group Pop. (2)	Ten year group Pop. (3)	k^4 (4)	k (5)
5 - 9	P_{5-9}	P_{5-14}		
10-14	P_{10-14}			
15-19	P_{15-19}	P_{15-24}	$(P_{5-14})/(P_{15-24})=k_1^4$	k_1
20-24	P_{20-24}			k_1
25-29	P_{25-29}	P_{25-34}	$(P_{15-24})/(P_{25-34})=k_2^4$	k_2
30-34	P_{30-34}			k_2
35-39	P_{35-39}	P_{35-44}	$(P_{25-34})/(P_{35-44})=k_3^4$	k_3
40-44	P_{40-44}			k_3

$k_1 = [(P_{5-14})/(P_{15-24})]^{\frac{1}{4}}$; $k_2 = [(P_{15-24})/(P_{25-34})]^{\frac{1}{4}}$ ...

.../107

Then we obtain the population in five year age groups as follows:

$$P_{15-19} = \frac{k_1(P_{15-24})}{1+k_1}$$

$$P_{20-24} = \frac{P_{15-24}}{1+k_1}$$

$$P_{25-29} = \frac{k_2(P_{25-34})}{1+k_2}$$

$$P_{30-34} = \frac{P_{25-34}}{1+k_2} \qquad \text{etc}$$

Thus we can obtain smoothed values only for the second ten year group to the last but one ten year group.

Let us obtain the values for male population aged 25-29 years in Tanzania by using the above 2 formulae: We need population aged 15-24, 25-34 and 35-44 which are respectively: $P_{15-24} = P_{15-19} + P_{20-24}$

$= 841340 + 586580$

$= 1427920$

$P_{25-34} = P_{25-29} + P_{30-34} = 610325 + 457537 = 1067862$ and

$P_{35-44} = P_{35-39} + P_{40-44} = 439515 + 321487 = 761002$

By Newton's halfing formula, $P_{25-29} = \frac{1}{16}\left[P_{15-24} + 8\,P_{25-34} - P_{35-44}\right]$

$= \frac{1}{16}\left[1427920 + 8(1067862) - 761002\right] = 575613$

By Carrier - Farrag formula, first we obtain

$k_2^4 = P_{15-24}/P_{25-34} = 1427920/1067862 = 1.337$

Hence $k_2 = 1.075$

Thus $P_{25-29} = \frac{k_2\ P_{25-34}}{1+k_2} = \frac{1.075(1067862)}{2.075} = 553406$

.../108

7.3 OSCULATORY INTERPOLATION

Osculatory interpolation formulas are another class of interpolation formulas designed for securing interpolated results which have a high degree of smoothness. They are particularly designed for use with rough data. Two formulas are presented here. These are the Karup-King Third-Difference Formula and the Sprague-Fifth-Difference Formula. Interpolation can be expressed in linear compound form, i.e., in terms of coefficients or multipliers that are applied to the given data. In this way, we only have to select the method of interpolation and know how to use the multipliers. We do not need to be familiar with the formula itself, nor with the mathematical derivation of the multipliers. Sets of multipliers are presented for the use of the two formulas above. Two types of formulas - one for interval and another for grouped data are presented in the following two tables in the next two pages.

Interval Data

The Karup-King formula is applied to four points while the Sprague formula is applied to six points. We may illustrate the application of the multipliers by interpolating survivors to single ages between ages 15 and 20 of the North Model life table level 11 for Females.

Survivors to age x out of 100,000 at age 0

Age	l_x (Survivors)
10	73775
15	71883
20	69917
25	67693
30	65200
35	62418

First the Karup-King method

The general form of the equation is:

$$N_{2+x} = m_1 N_{1.0} + m_2 N_{2.0} + m_3 N_{3.0} + m_4 N_{4.0}$$

where x is a fraction between 0 and 1.

$N_{1.0}$, $N_{2.0}$, $N_{3.0}$ and $N_{4.0}$ represent four given values

m_1, m_2, m_3 and m_4 are the four multipliers associated with the four given points.

Table [illegible].—Interpolation Coefficients Based on the Karup-King Formula

[The Karup-King formula is a four-term third-difference osculatory formula. It maintains the given values. Given points or groups must be equally spaced]

A. FOR INTERPOLATION BETWEEN GIVEN POINTS AT INTERVALS OF 0.2

Interpolated point	Coefficients to be applied to— $N_{1.0}$	$N_{2.0}$	$N_{3.0}$	$N_{4.0}$
	First interval			
$N_{1.0}$	+1.000	.000	.000	.000
$N_{1.2}$	+.656	+.552	-.272	+.064
$N_{1.4}$	+.408	+.856	-.336	+.072
$N_{1.6}$	+.232	+.984	-.264	+.048
$N_{1.8}$	+.104	+1.008	-.128	+.016
	Middle interval			
$N_{2.0}$	.000	+1.000	.000	.000
$N_{2.2}$	-.064	+.912	+.168	-.016
$N_{2.4}$	-.072	+.696	+.424	-.048
$N_{2.6}$	-.048	+.424	+.696	-.072
$N_{2.8}$	-.016	+.168	+.912	-.064
	Last interval			
$N_{3.0}$	.000	.000	+1.000	.000
$N_{3.2}$	+.016	-.128	+1.008	+.104
$N_{3.4}$	+.048	-.264	+.984	+.232
$N_{3.6}$	+.072	-.336	+.856	+.408
$N_{3.8}$	+.064	-.272	+.552	+.656
$N_{4.0}$	.000	.000	.000	+1.000

B. FOR SUBDIVISION OF GROUPS INTO FIFTHS

Interpolated subgroup	Coefficients to be applied to— G_1	G_2	G_3
	First panel		
First fifth of G_1	+.344	-.208	+.064
Second fifth of G_1	+.248	-.056	+.008
Third fifth of G_1	+.176	+.048	-.024
Fourth fifth of G_1	+.128	+.104	-.032
Last fifth of G_1	+.104	+.112	-.016
	Middle panel		
First fifth of G_2	+.064	+.152	-.016
Second fifth of G_2	+.008	+.224	-.032
Third fifth of G_2	-.024	+.248	-.024
Fourth fifth of G_2	-.032	+.224	+.008
Last fifth of G_2	-.016	+.152	+.064

B. FOR SUBDIVISION OF GROUPS INTO FIFTHS—Continued

Interpolated subgroup	Coefficients to be applied to— G_1	G_2	G_3
	Last panel		
First fifth of G_3	-.016	+.112	+.104
Second fifth of G_3	-.032	+.104	+.128
Third fifth of G_3	-.024	+.048	+.176
Fourth fifth of G_3	+.008	-.056	+.248
Last fifth of G_3	+.064	-.208	+.344

C. FOR SUBDIVISION OF GROUPS INTO TENTHS OR HALVES

Interpolated subgroup	Coefficients to be applied to— G_1	G_2	G_3
First tenth of G_2	+.0405	+.0640	-.0045
Second tenth of G_2	+.0235	+.0880	-.0115
Third tenth of G_2	+.0095	+.1060	-.0155
Fourth tenth of G_2	-.0015	+.1180	-.0165
Fifth tenth of G_2	-.0095	+.1240	-.0145
Sum of coefficients for first five-tenths = coefficients for first half of G_2	+.0625	+.5000	-.0625
Sixth tenth of G_2	-.0145	+.1240	-.0095
Seventh tenth of G_2	-.0165	+.1180	-.0015
Eighth tenth of G_2	-.0155	+.1060	+.0095
Ninth tenth of G_2	-.0115	+.0880	+.0235
Last tenth of G_2	-.0045	+.0640	+.0405
Sum of coefficients for last five-tenths = coefficients for last half of G_2	-.0625	+.5000	+.0625

Table – Interpolation Coefficients Based on the Sprague Formula

[The Sprague formula is a six-term fifth-difference osculatory formula. It maintains the given values. Given points or groups must be equally spaced]

A. FOR INTERPOLATION BETWEEN GIVEN POINTS AT INTERVALS OF 0.2

Interpolated point	Coefficients to be applied to-- $N_{1.0}$	$N_{2.0}$	$N_{3.0}$	$N_{4.0}$	$N_{5.0}$	$N_{6.0}$
	First interval					
$N_{1.0}$	+1.0000	.0000	.0000	.0000	.0000	
$N_{1.2}$	+.6384	+.6384	-.4256	+.1824	-.0336	
$N_{1.4}$	+.3744	+.9984	-.5616	+.2304	-.0416	
$N_{1.6}$	+.1904	+1.1424	-.4896	+.1904	-.0336	
$N_{1.8}$	+.0704	+1.1264	-.2816	+.1024	-.0176	
	Next-to-first interval					
$N_{2.0}$	.0000	+1.0000	.0000	.0000	.0000	
$N_{2.2}$	-.0336	+.8064	+.3024	-.0896	+.0144	
$N_{2.4}$	-.0416	+.5824	+.5824	-.1456	+.0224	
$N_{2.6}$	-.0336	+.3584	+.8064	-.1536	+.0224	
$N_{2.8}$	-.0176	+.1584	+.9504	-.1056	+.0144	
	Middle interval					
$N_{3.0}$	.0000	.0000	+1.0000	.0000	.0000	.0000
$N_{3.2}$	+.0128	-.0976	+.9344	+.1744	-.0256	+.0016
$N_{3.4}$	+.0144	-.1136	+.7264	+.4384	-.0736	+.0080
$N_{3.6}$	+.0080	-.0736	+.4384	+.7264	-.1136	+.0144
$N_{3.8}$	+.0016	-.0256	+.1744	+.9344	-.0976	+.0128
	Next-to-last interval					
$N_{4.0}$		.0000	.0000	+1.0000	.0000	.0000
$N_{4.2}$		+.0144	-.1056	+.9504	+.1584	-.0176
$N_{4.4}$		+.0224	-.1536	+.8064	+.3584	-.0336
$N_{4.6}$		+.0224	-.1456	+.5824	+.5824	-.0416
$N_{4.8}$		+.0144	-.0896	+.3024	+.8064	-.0336
	Last interval					
$N_{5.0}$		.0000	.0000	.0000	+1.0000	.0000
$N_{5.2}$		-.0176	+.1024	-.2816	+1.1264	+.0704
$N_{5.4}$		-.0336	+.1904	-.4896	+1.1424	+.1904
$N_{5.6}$		-.0416	+.2304	-.5616	+.9984	+.3744
$N_{5.8}$		-.0336	+.1824	-.4256	+.6384	+.6384
$N_{6.0}$		.0000	.0000	.0000	.0000	+1.0000

B. FOR SUBDIVISION OF GROUPS INTO FIFTHS

Interpolated subgroup	Coefficients to be applied to-- G_1	G_2	G_3	G_4	G_5
	First panel				
First fifth of G_1	+.3616	-.2768	+.1488	-.0336	
Second fifth of G_1	+.2640	-.0960	+.0400	-.0080	
Third fifth of G_1	+.1840	+.0400	-.0320	+.0080	
Fourth fifth of G_1	+.1200	+.1360	-.0720	+.0160	
Last fifth of G_1	+.0704	+.1968	-.0848	+.0176	
	Next-to-first panel				
First fifth of G_2	+.0336	+.2272	-.0752	+.0144	
Second fifth of G_2	+.0064	+.2320	-.0480	+.0096	
Third fifth of G_2	-.0080	+.2160	-.0080	.0000	
Fourth fifth of G_2	-.0144	+.1840	+.0400	-.0096	
Last fifth of G_2	-.0176	+.1408	+.0912	-.0144	

B. FOR SUBDIVISION OF GROUPS INTO FIFTHS--Continued

Interpolated subgroup	Coefficients to be applied to-- G_1	G_2	G_3	G_4	G_5
	Middle panel				
First fifth of G_3	-.0128	+.0848	+.1504	-.0240	+.0016
Second fifth of G_3	-.0016	+.0144	+.2224	-.0416	+.0064
Third fifth of G_3	+.0064	-.0336	+.2544	-.0336	+.0064
Fourth fifth of G_3	+.0064	-.0416	+.2224	+.0144	-.0016
Last fifth of G_3	+.0016	-.0240	+.1504	+.0848	-.0128
	Next-to-last panel				
First fifth of G_4		-.0144	+.0912	+.1408	-.0176
Second fifth of G_4		-.0096	+.0400	+.1840	-.0144
Third fifth of G_4		.0000	-.0080	+.2160	-.0080
Fourth fifth of G_4		+.0096	-.0480	+.2320	+.0064
Last fifth of G_4		+.0144	-.0752	+.2272	+.0336
	Last panel				
First fifth of G_5		+.0176	-.0848	+.1968	+.0704
Second fifth of G_5		+.0160	-.0720	+.1360	+.1200
Third fifth of G_5		+.0080	-.0320	+.0400	+.1840
Fourth fifth of G_5		-.0080	+.0400	-.0960	+.2640
Last fifth of G_5		-.0336	+.1488	-.2768	+.3616

C. FOR SUBDIVISION OF GROUPS INTO TENTHS OR HALVES

Interpolated subgroup	Coefficients to be applied to-- G_1	G_2	G_3	G_4	G_5
First tenth of G_3	-.0076	+.0510	+.0660	-.0096	+.0002
Second tenth of G_3	-.0052	+.0338	+.0844	-.0144	+.0014
Third tenth of G_3	-.0022	+.0154	+.1036	-.0195	+.0027
Fourth tenth of G_3	+.0006	-.0010	+.1188	-.0221	+.0037
Fifth tenth of G_3	+.0027	-.0133	+.1272	-.0203	+.0037
Sum of coefficients for first five-tenths = coefficients for first half of G_3	-.0117	+.0859	+.5000	-.0859	+.0117
Sixth tenth of G_3	+.0037	-.0203	+.1272	-.0133	+.0027
Seventh tenth of G_3	+.0037	-.0221	+.1188	-.0010	+.0006
Eighth tenth of G_3	+.0027	-.0195	+.1036	+.0154	-.0022
Ninth tenth of G_3	+.0014	-.0144	+.0844	+.0338	-.0052
Last tenth of G_3	+.0002	-.0096	+.0660	+.0510	-.0076
Sum of coefficients for last five-tenths = coefficients for second half of G_3	+.0117	-.0859	+.5000	+.0859	-.0117

We want to estimate the l_x values
i.e. l_{16}, l_{17}, l_{18} and l_{19}

The four given values in this case are

l_{10}, l_{15}, l_{20} and l_{25}

Note that the range l_{15} to l_{20} is the middle interval. The table has intervals which are first interval, middle interval, and last interval. Different multipliers are applicable to interpolated points of different intevals.

Also note that

l_{16} is a value of 0.2 units of the way from l_{15} to l_{20}.
Similarly l_{17} is a value of 0.4, l_{18} is a value of 0.6 and l_{19} is a value of 0.8.

Hence we have $N_{2.0}$ for l_{15}, $N_{2.2}$ for l_{16}, $N_{2.4}$ for l_{17}, $N_{2.6}$ for l_{18} and $N_{2.8}$ for l_{19}.

Hence

$$l_{16} = m_1 l_{10} + m_2 l_{15} + m_3 l_{20} + m_4 l_{25}$$

Selecting the multipliers for l_{16} in the middle interval we have

$$l_{16} = -0.064\ l_{10} + 0.912\ l_{15} + 0.168\ l_{20} - 0.016\ l_{25}$$

Substituting the values for l_{10}, l_{15}, l_{20} and l_{25} we have

$$l_{16} = -0.064(73775) + 0.912(71883) + 0.168(69917) + 0.016(67693)$$
$$= 77303.352 - 5804.688 \doteq 71499$$

Similarly for l_{17} where

$$l_{17} = -0.072(73775) + 0.696(71883) + 0.4224(69917) - 0.048(67693)$$

Note that the multipliers are now different. These correspond to $N_{2.4}$ the symbol for l_{17}. The same procedure can be done for the other ages, i.e. l_{18} and l_{19}.

The application of the Sprague multipliers is the same except that these use six terms. Let us demonstrate this using the same example above for l_{16}.

$$N_{2+x} = m_1 N_{1.0} + m_2 N_{2.0} + m_3 N_{3.0} + m_4 N_{4.0} + m_5 N_{5.0} + m_6 N_{6.0}$$

i.e. $l_{16} = m_1 l_{10} + m_2 l_{15} + m_3 l_{20} + m_4 l_{25} + m_5 l_{30} + m_6 l_{35}$

Using the multipliers from the tables (note: for l_{15} to l_{20}; the appropriate values are in the next to first interval) and the corresponding l_x values we have

.../110

$$l_{16} = -0.0336(73775) + 0.8064(71883) + 0.3024\ (69917) - 0.0896(67693) + 0.0144(65200) = 71504.$$

The same can be done for the other ages.

Grouped Data

So far we have been dealing with interpolation of point data. We now turn to interpolation of grouped or area data. Interpolation of grouped data serves many purposes. The most common use is the estimation of data in finer detail than is available in published data, e.g. for estimating single year figures from published data by five-year age groups. Another purpose is for smoothing or graduation of data. Data given in single years can be grouped into five-year groups, then interpolated to obtain smoothed esimates by single years of age. Still another purpose may be to redistribute data given in one grouping to another required grouping. It should be noted that the methods described assumed that the pattern of distribution of grouped data is a valid indication of the pattern of the distribution within groups.

The simplest and rather most commonly used method of subdividing group data employes the assumption that the data are evenly (or uniformly, or rectangularly) distributed within the interval to be subdivided. The assumption is that the values of the part are all equal. These are then derived by dividing the total of the interval by the number of parts desired. This approach is a useful basis for obtaining quick and rough estimates of detailed categories under many different circumstances in demographic analysis.

Graphs are useful for deriving rough estimates for subdivisions of grouped data particularly when the groups are unevenly spaced. One procedure for graphic interpolation is to draw smooth lines through the mid-points of the histograms and then obtain the interpolated results by reading off the values of the ordinates along the curve that are at the desired abscissas.

The Waring's procedure can also be used for interpolation for grouped data. However, Osculatory interpolation is preferred for interpolating demographic data. Again only two methods are demonstrated. These are the Karup-King Formula and the Sprague Formula. Their application using the multipliers is demonstrated rather than the actual derivation of the formula. Again we refer to the respective tables provided

Let us use the 1978 male population in Tanzania given in Table 6

We first apply the Karup-King multipliers which use the Third-Difference formula. We want to estimate the population for single years in the age group 20-24 using age groups 15-19, 20-24 and 25-29. The age group 20-24 is the middle panel. So we use the multipliers of the middle panel. Age 20 is the first fifth, 21 the second fifth, 22 the third fifth, 23, the fourth fifth and 24 the last fifth.

For example the multipliers for age 20 which is the first fifth are given as:

	G_1	G_2	G_3
First fifth of G_2	0.064	0.152	-0.016

G_1 , G_2 and G_3 stand respectively for the population in age group 15-19 20 - 24, and 25 - 29

Let P_{20} be the estimated population for age 20

Then

$$P_{20} = 0.064\ N_1 + 0.152\ N_2 - 0.016\ N_3$$

where N_1, N_2 and N_3 are the respective populations for age groups 15-19, 20-24 and 25-29.

Hence

$$P_{20} = 0.064\ (841340) + 0.152\ (586580) - 0.016\ (610325) = 133241$$

Similarly

$$P_{21} = 0.008\ (841340) + 0.224\ (586580) - 0.032\ (610325)$$
$$= 118594$$

The same procedure can be applied to the Sprague formula using the multipliers provided in the table. The formula for the population aged 20 years is

$$P_{20} = m_1N_1 + m_2N_2 + m_3N_3 + m_4N_4 + m_5N_5 + m_6N_6$$

where

P_{20} = population for age 20

m_i = are the multipliers for the respective groups

N_i = the population for the respective age groups starting with age group 15 - 19.

Hence

$$P_{20} = -.0128(841340) + .0848(586580) + .1504(610325) -.0240(457537)$$
$$+ .0016(439515)$$
$$= 120488$$

7.4 <u>Curve Fitting</u>

Demographic data may follow a typical trend or pattern that can be represented empirically by some mathematical equation. Curve fitting consists of finding a suitable equation to represent that trend or pattern. Curves to be fitted may be polynomials, exponential equations, trigonometric equations (useful for data that have periodic fluctuations or seasonal

patterns), or still other curves. The aim may be to fit a curve to the data in an approximate fashion, in which case crude methods, e.g. graphic or moving averages, may be suitable. At times it may be necessary to fit a curve using sophisticated methods, e.g. moments or least squares. The suitability of the fitted curve for interpolation or extrapolation would depend on the nature of the given data, on the choice of curve, and on the goodness of fit.

The method of least squares is one of the most popular methods applied in curve fitting. This method minimizes the sum of the squares of the differences between the observed points and the points calculated from the fitted curve.(The expected values)

i.e. $\Sigma(Y - \hat{Y})^2$ = minimum, where Y is the observed value and $\hat{Y}$ is the corresponding value from the fitted curve.

For example if a straight line is to be fitted to a series of data, we have the equation representing a straight line as

$$Y = A + BX$$

where A is the intercept on the Y-axis and B is the slope.

All we need is to estimate the two coefficients A and B using the observed values of X and Y, say given for N points.
By the method of least squares we obtain the equations (known as the Normal Equations)

$$\Sigma Y = NA + B\Sigma X$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2$$

By obtaining ΣY, ΣX, ΣXY and ΣX^2 and solving for A and B the equation of the line can be obtained. Higher degree curves can be fitted in a similar way. For example a second-degree polynomial is of the form

$$Y = A + BX + CX^2$$

The normal equations for this are

$$\Sigma Y = NA + B\Sigma X + C\Sigma X^2$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2 + C\Sigma X^3$$

$$\Sigma X^2Y = A\Sigma X^2 + B\Sigma X^3 + C\Sigma X^4$$

All we need to solve for A, B and C are: ΣY, ΣX, ΣXY, ΣX^2Y, ΣX^2, ΣX^3 and ΣX^4

Exponential functions are another class of mathematical equations useful in interpolation and extrapolation as well as curve fitting for a series of data. An exponential function is one in which one or more of the variables is expressed as a power of some parameter or constant in a formula.

e.g. $Y = A^X$ is an exponential function because X is a power of the

parameter A. But $Y = X^A$ is not an exponential function because the variable X is not a power of parameter.

Exponential functions take many forms. A very general form is the power function

$$Y = AB^X$$

This is also known as the growth curve. The geometric and exponential growth curves are special cases.

This equation can be written as : $\log Y = \log A + X \log B$, i.e. $y = a + bX$,, i.e. a linear form between $y = \log Y$ and X and hence the method of least squares shown for fitting a linear equation can be applied. There are other growth curves, some of which are modifications of the general exponential equation

$$Y = AB^X$$

e.g. (1) $Y = K + AB^X$. This is the modified exponential curve

(2) The Gompertz Curve is a popular modified exponential equation given by:

$$y = ka^{B^X}$$

which can be reduced to

$$\log y = \log k + (\log a)B^X = K + AB^X \text{ where } K = \log k \text{ and } A = \log a.$$

(3) Yet another popular type is the logistic curve also known as the Pearl-Reed Curve given by

$$\frac{1}{Y} = K + AB^X$$

and can be modified to

$$Y = \frac{1}{K + AB^x} = \frac{c}{1 + e^{a+bx}}$$

This is generally applied to population growth with C as the assymptote, i.e., the upper limit for the growth of the population will be C.

Let us illustrate the fitting of a straight line and exponential curve by using the population figures for Tanzania between 1948 and 1967.

The population in 1948, 1957 and 1967 in Tanzania were respectively:

Year, t	1948	1957	1967
Population, y	7744600	9084100	12313054
(t-1957) = x	-9	0	10
ln y = Y	15.862	16.022	16.326

The transformed variable x is taken because it reduces the magnitude of the values and hence reduces calculation. We could have taken a scale also,

had the intervals been equal or such as not to increase the computational load.

For fitting the straight line the normal equations are:

Sum of y = na+ b(Sum of x) and
Sum of xy = a(Sum of x)+ b(Sum of x^2)

Substituting the values of n, sum of x, sum of y, sum of x^2 and sum of xy, we get,

$$29141754 = 3a + b$$

$$53429140 = a + 181\,b$$

Hence a = 9633263 and b = 241966

For fitting the exponential curve we deal with log y = Y and proceed exactly as above. The normal equations are:

$$48.210 = 3a + b$$

$$20.502 = a + 181\,b$$

Solving we get a = 16.062 and b = .025.

7.5 DEMOGRAPHIC MODELS

A model may be defined as a conceptualization or representation of a phenomenon or process. Demographic models represent generalized illustration of a demographic process or phenomenon and it could be of various types. For example, statement that a population is closed to migration produces the concept of a population which changes only through the natural process of births and deaths. Mathematical functions or formulas like the relation between male and female births in a population for a given sex proportion at birth or more complicated functions depicting the curve of fertility or mortality are examples. Models also could be presented in tabular or graphic forms as in the case of a schedule of age specific fertility corresponding to a given gross reproduction rate.

Models are used to represent some process as fully as possible especially when there is insufficiency of data. It may also be used to remove some of the errors inherent in data to make the data more useable.

Increasing use is being made of model building in demography. The life table is one such. More complex models link together component models of fertility, mortality and migration. Stable population is yet another well known model.

Life Tables

Mathematical functions like those of Gompertz and Makeham have been used in the study of mortality but inaccuracy in data would make the use of mathematical functions difficult and risky. Comparison with other experiences would seem better as in the analogy method. The difficulty here would be the closeness of the two experiences in respect of the phenomenon being investigated, especially when the situation on hand is not very clear.

Model life tables

In the study of mortality in Europe, Dudley Kirk and associates observed that the mortality rates at adjacent ages are highly correlated. Thus if some strand of information is available on the mortality, then this correlation could be utilized to obtain the entire mortality experience. This idea was exploited in 1955 by the United Nations who related mortality at various ages through a chain regression equation starting with infant mortality and early childhood mortality. From infant mortality, through the regression equation, the estimate of child mortality is derived which then is used sequentially to obtain mortality at other adjacent ages. Even though it was a very useful first step in the construction of a life table, because then one would need only an estimate of infant mortality, three drawbacks will have to be kept in mind. The first is that there could be statistical bias introduced by the chain like series of regression equation. Secondly, the life tables used to derive the regression parameters may not be representative of the mortality experience on hand and thirdly, the validity of considering infant mortality as an index of mortality especially in view of the heterogeneity of age-sex patterns of mortality has been questioned. Some of these drawbacks have been removed by the development of the Regional model life tables by Coale and associates at Princeton University. Two other model systems developed are those due to Brass and Lederman. Recently UN has prepared revised model life tables.

In situations where mortality pattern is not available, it is recommended to use the Coale-Demeny West model because it is based on the experience of a large number of countries and may not be too far from a normal pattern of mortality. However, if some information is available, like the excessive child mortality as in some African countries in comparison with infant mortality, then the North pattern may be more appropriate.

The model life table has become a very useful tool especially in the preparation of population projections and for studies on mortality in general. Of great value in demography has been the development of stable population models based on model life tables. A sample page from UN Manual IV showing the life tables at level 11 (West model) is given in Table in the next page.

Stable population models

In a population with constant fertility and mortality conditions, which is closed to migration, the age-sex structure tends to a form called the stable age distribution after these conditions have been operative for a sufficiently long time. In such a population the percentage age distribution remains constant and the growth rate becomes constant. The equation to a stable population with given fertility-mortality rates is given by the equation:

$C(x) = b\, e^{rx}\, l_x$, where b is the birth rate, r the growth rate and l_x is the probability of surviving from birth to exact age x. (c(x) is the proportion of persons at age x). When we consider population in an age group, say x to x+4, then the corresponding equation becomes:

.../116

TABLE 1. "WEST" MODEL LIFE TABLES ARRANGED BY LEVEL OF MORTALITY (*continued*)

LEVEL 11

Age x	l_x	$_nm_x$	$_nq_x$	$_nL_x$	$\frac{_5L_{x+5}}{_5L_x}$	T_x	$\mathring{e}_x$
			Females				
0	100,000	.1615	.1461	90,502	.8219[a]	4,500,000	45.00
1	85,388	.0250	.0937	320,442	.9288[b]	4,409,498	51.64
5	77,389	.0055	.0272	381,683	.9758	4,089,056	52.84
10	75,285	.0043	.0212	372,430	.9752	3,707,373	49.25
15	73,687	.0058	.0284	363,207	.9679	3,334,942	45.26
20	71,596	.0073	.0360	351,543	.9618	2,971,735	41.51
25	69,022	.0083	.0405	338,115	.9569	2,620,192	37.96
30	66,224	.0094	.0459	323,525	.9516	2,282,077	34.46
35	63,186	.0105	.0510	307,872	.9465	1,958,552	31.00
40	59,963	.0116	.0562	291,388	.9402	1,650,680	27.53
45	56,592	.0131	.06[illegible]6	273,969	.9267	1,359,291	24.02
50	52,996	.0175	.0837	253,884	.9039	1,085,322	20.48
55	48,558	.0232	.1095	229,493	.8669	831,439	17.12
60	43,239	.0347	.1596	198,946	.8127	601,946	13.92
65	36,339	.0495	.2203	161,682	.7367	403,000	11.09
70	28,333	.0758	.3184	119,112	.6304	241,319	8.52
75	19,311	.1144	.4448	75,083	.3856	122,207	6.33
80	10,722	.2275	—	47,125		47,124	4.40
			Males				
0	100,000	.1940	.1717	88,499	.7983[a]	4,211,576	42.12
1	82,835	.0252	.0944	310,632	.9274[b]	4,123,076	49.78
5	75,015	.0053	.0262	370,157	.9773	3,812,445	50.82
10	73,048	.0038	.0190	361,763	.9771	3,442,288	47.12
15	71,657	.0054	.0268	353,478	.9677	3,080,525	42.99
20	69,734	.0078	.0380	342,042	.9601	2,727,047	39.11
25	67,083	.0086	.0419	328,380	.9550	2,385,005	35.55
30	64,269	.0099	.0482	313,607	.9476	2,056,625	32.00
35	61,173	.0117	.0569	297,166	.9368	1,743,018	28.49
40	57,693	.0145	.0698	278,395	.9231	1,445,852	25.06
45	53,665	.0177	.0845	256,985	.9032	1,167,457	21.76
50	49,129	.0233	.1102	232,105	.8750	910,472	18.53
55	43,713	.0305	.1416	203,093	.8337	678,366	15.52
60	37,524	.0432	.1951	169,318	.7743	475,273	12.67
65	30,203	.0607	.2636	131,109	.6951	305,955	10.13
70	22,241	.0881	.3610	91,133	.5879	174,847	7.86
75	14,213	.1306	.4922	53,575	.3600[c]	83,714	5.89
80	7,217	.2395	—	30,139		30,139	4.18

[a] Proportion surviving from birth to 0–4.
[b] $_5L_5/_5L_0$.
[c] T_{80}/T_{75}.

$c(x, x+4) = b\, e^{r(x+2.5)}\, L_{x, x+4}$, where $L_{x, x+4}$ is the life table population aged x to x+ 4 and $c(x, x+4)$ is the population aged x to x+ 4 and b and r have the same conotation as in the previous equation.

Stable population models have been calculated based on the UN model life tables by the CELADE and by Coale and Demeny based on the Regional model life tables. Based on the new life tables prepared by UN sets of stable populations have been produced. Stable models based on Brass life tables are given by Carrier-Hobcraft.

The stable population model is a powerful tool in the evaluation of data and has been successfully employed to estimate demographic parameters in countries with defective and incomplete data. Some of these will be discussed in later sections.

Allied to the concept of stability is the concept of quasi stability which is defined when fertility remains constant for a very long time but mortality might be falling but not drastically. In most of the countries in the developing part of the world, this might be the situation existing and hence quasi stable population theory and models become appropriate.

A sample page from UN Manual IV showing the female stable population (West model) for level 11 and various growth rates is given in Table in next page.

8. ANALYSIS OF MORTALITY DATA

Introduction

In most of the developing countries information on deaths is the weakest. Either direct information on mortality are non-existent or are grossly deficient. Hence recourse will have to be taken to indirect methods to estimate mortality.

Two such approaches are (1) based on age-sex distribution and (2) child survival ratios.

8.1 ESTIMATION OF MORTALITY FROM AGE STRUCTURE

The age-sex distribution of a population is the end result of fertility, mortality and migration. Thus, if a population can be considered to be closed to migration, the age-sex distribution reflects fertility - mortality conditions.

In situations where fertility/mortality have remained constant for a very long period of time, it can be shown that the age-sex distribution stabilises i.e. has a constant rate of growth for each age sex segment and hence the percentage in any age group will remain constant and sex ratio will not fluctuate. Even when mortality changes, but only gradually, and fertility remains constant, we obtain an age-sex distribution which is not too different from that resulting from constant mortality.

TABLE II. "WEST" MODEL STABLE POPULATIONS ARRANGED BY LEVEL OF MORTALITY (*continued*)

LEVEL 11

Females (${}^{o}e_0$ = 45.00 *years*)

	Annual rate of increase												
	−.010	*−.005*	*.000*	*.005*	*.010*	*.015*	*.020*	*.025*	*.030*	*.035*	*.040*	*.045*	*.050*
					Proportion in age interval								
Age interval													
Under 1	.0142	.0170	.0201	.0235	.0272	.0311	.0352	.0396	.0440	.0487	.0534	.0582	.0631
1-4	.0516	.0610	.0712	.0822	.0939	.1061	.1187	.1316	.1447	.1579	.1711	.1842	[illegible]
5-9	.0643	.0743	.0848	.0957	.1069	.1181	.1292	.1401	.1506	.1606	.1702	.1792	.1875
10-14	.0660	.0743	.0828	.0911	.0992	.1069	.1141	.1206	.1265	.1316	.1360	.1396	.1425
15-19	.0676	.0743	.0807	.0867	.0920	.0967	.1006	.1038	.1061	.1077	.1086	.1087	.1082
20-24	.0688	.0738	.0781	.0818	.0847	.0868	.0881	.0887	.0884	.0875	.0860	.0840	.0816
25-29	.0696	.0727	.0751	.0767	.0775	.0775	.0767	.0753	.0732	.0707	.0677	.0645	.0611
30-34	.0700	.0714	.0719	.0716	.0706	.0688	.0664	.0635	.0603	.0568	.0531	.0493	.0455
35-39	.0700	.0696	.0684	.0665	.0639	.0607	.0572	.0534	.0494	.0453	.0413	.0375	.0337
40-44	.0697	.0676	.0648	.0614	.0575	.0533	.0490	.0446	.0402	.0360	.0320	.0283	.0249
45-49	.0689	.0651	.0609	.0563	.0514	.0465	.0417	.0370	.0326	.0284	.0247	.0213	.0182
50-54	.0671	.0619	.0564	.0508	.0453	.0400	.0349	.0302	.0260	.0221	.0187	.0157	.0131
55-59	.0637	.0574	.0510	.0448	.0390	.0335	.0286	.0241	.0202	.0168	.0138	.0114	.0093
60-64	.0581	.0510	.0442	.0379	.0321	.0270	.0224	.0185	.0151	.0122	.0098	.0079	.0062
65-69	.0496	.0425	.0359	.0300	.0248	.0203	.0165	.0132	.0105	.0083	.0065	.0051	.0040
70-74	.0384	.0321	.0265	.0216	.0174	.0139	.0110	.0086	.0067	.0052	.0039	.0030	.0023
75-79	.0255	.0207	.0167	.0133	.0104	.0081	.0063	.0048	.0036	.0027	.0020	.0015	.0011
80+	.0170	.0134	.0105	.0081	.0062	.0047	.0035	.0026	.0019	.0014	.0010	.0007	.0005
					Proportion under given age								
Age													
1	.0142	.0170	.0201	.0235	.0272	.0311	.0352	.0396	.0440	.0487	.0534	.0582	.0631
5	.0658	.0780	.0913	.1057	.1210	.1371	.1539	.1711	.1887	.2065	.2245	.2424	.2602
10	.1301	.1523	.1761	.2014	.2279	.2552	.2831	.3112	.3393	.3672	.3947	.4215	.4477
15	.1961	.2266	.2589	.2926	.3271	.3621	.3971	.4318	.4658	.4988	.5306	.5611	.5902
20	.2637	.3009	.3396	.3792	.4191	.4588	.4978	.5356	.5719	.6065	.6392	.6698	.6984
25	.3325	.3747	.4177	.4610	.5038	.5457	.5859	.6242	.6604	.6940	.7252	.7539	.7800
30	.4021	.4474	.4929	.5377	.5814	.6231	.6626	.6995	.7336	.7647	.7930	.8184	.8411
35	.4721	.5188	.5648	.6094	.6519	.6919	.7290	.7630	.7938	.8215	.8460	.8677	.8867
40	.5421	.5884	.6332	.6758	.7158	.7527	.7862	.8164	.8432	.8668	.8874	.9052	.9204
45	.6117	.6559	.6979	.7372	.7733	.8060	.8352	.8610	.8835	.9028	.9194	.9335	.9453
50	.6806	.7211	.7588	.7934	.8247	.8525	.8769	.8980	.9160	.9313	.9441	.9547	.9635
55	.7476	.7829	.8152	.8443	.8700	.8925	.9118	.9282	.9420	.9534	.9628	.9705	.9766
60	.8114	.8403	.8662	.8891	.9090	.9260	.9404	.9523	.9622	.9702	.9766	.9818	.9859
65	.8695	.8913	.9104	.9270	.9411	.9530	.9628	.9708	.9772	.9824	.9865	.9897	.9921
					Parameters of stable populations								
Birth rate	.0156	.0188	.0222	.0260	.0302	.0346	.0393	.0443	.0494	.0547	.0602	.0658	.0715
Death rate	.0256	.0238	.0222	.0210	.0202	.0196	.0193	.0193	.0194	.0197	.0202	.0208	.0215
GRR (27)	1.13	1.29	1.48	1.69	1.92	2.19	2.50	2.84	3.23	3.66	4.16	4.71	5.33
GRR (29)	1.13	1.30	1.50	1.73	2.00	2.30	2.64	3.03	3.47	3.98	4.55	5.20	5.94
GRR (31)	1.12	1.31	1.53	1.78	2.07	2.41	2.79	3.24	3.75	4.34	5.01	5.78	6.68
GRR (33)	1.12	1.32	1.56	1.83	2.15	2.53	2.97	3.47	4.07	4.76	5.56	6.48	7.56
Average age	37.6	35.2	32.9	30.6	28.5	26.4	24.6	22.8	21.2	19.7	18.4	17.2	16.1
Births/population 15-44	.038	.044	.051	.059	.068	.078	.090	.103	.118	.135	.155	.177	.201

But in the latter case, the age sex distribution will not have constant growth rate. The resulting distribution is called 'quasi stable'. The quasi stable population resembles a 'stable' population with recent fertility, mortality and growth rate as parameters.

In developing countries, fertility can be assumed to have remained high and constant for the past several generations but that mortality must have improved. Thus their reported age sex distributions should be near a quasi stable population, excepting for distortions brought in by age-sex reporting errors and biases.

Several methods have been developed based on stability or quasi stability of population and because of the distortions in age-sex distribution, these methods use techniques which remove some of the biases by averaging or cumulation using several ages.

One such method is that due to Arriaga. In this method, age sex distribution is utilized to arrive at estimates of mortality and as a bye-product the birth rate also. Only it uses the reported age sex distribution in the first 7 or 9 age groups and the estimated rate of growth of population in a recent period.

The method is based on the mathematical relation between age distribution and birth and death rates. The stable age distribution has the equation as already indicated in section 7.5:

$C(x) = b\,e^{-rx}\,l_x$ for single year of ages and for five year age groups the equation becomes:

$$C(x, x+4) = b\ e^{-r(x+2.5)} L_{x,\ x+4}$$

This equation can be rewritten as:

$$\frac{C_{(x,\ x+4)}}{L_{x,\ x+4}} = b\ e^{-r(x+2.5)}$$

$$\text{or } \ln \frac{C(x,\ x+4)}{Lx,\ x+4} = \ln b - r(x+2.5) = A + Bx \quad \text{where,}$$

$A = \ln b - 2.5\,r$ and $B = -r$

Thus the relation between $\ln \frac{C(x,\ x+4)}{Lx,\ x+4}$ and x is linear with coefficients depending only on b and r. Hence if r is known, the method uses the fact that the slope of the line is - r and finds the values of Lx, x+4 from appropriate life table such that the resulting line has slope - r.

Because of irregularity in age data through reporting errors etc, we use 7 or 9 C(x, x+4) values starting with age 0 - 4 and by method of least squares identify the life table which fits in with known growth rate. As a bye-product, we can estimate b from the value A (intercept of the fitted line).

.../118

The normal equations obtained by method of least squares are:

$$\text{Sum of } \ln \frac{C(x,\ x+4)}{Lx,\ x+4} = nA + B.\ \text{Sum of } x$$

$$\text{and Sum of } x \ln \frac{C(x,\ x+4)}{Lx,\ x+4} = A.\ \text{Sum of } x + B.\ \text{Sum of squares of } x$$

Solving the 2 equations we get A and B. The life table which produces a value of B near - r is taken as applicable to the population.

It is advisable to perform the method for each sex separately. Nromally the level of mortality indicated for the sexes must be near each other. Wide fluctuations would indicate the problem of age reporting errors. If age reporting are affected in opposite ways for the sexes, then an average of the levels would be near the true value. Sometimes, depending on quality of age reporting, the data for one sex may indicate better estimates than the other.

In the application of the above method, either the actual numbers in the various ages or the percentages can be used. Also instead of the actual age groups, we can transform them into simple values by choosing an appropriate origin and scale. When an origin and scale is chosen, the values of r and b will have to be worked out in terms of A and B carefully. If an origin is chosen in the middle of the age range as in Table 19, then the sum of x will be zero and the normal equations will be simplified.

Table 19 ARRIAGE METHOD FOR ESTIMATING MORTALITY LEVEL, Tanzania
(Using North model level 11)

M a l e :

Age, x (1)	Pop. C(x, x+4) % (2)	Life table Lx, x+4 (3)	(4) = (2)/(3) x 10000	(5) = ln(4)	X (6)	(7) = (6)(5)
0 - 4	18.14	405198	.4477	-.8036	-3	2.4108
5 - 9	16.20	365955	.4427	-.8149	-2	1.6298
10-14	12.42	352283	.3526	-1.0424	-1	1.0424
15-19	9.80	342668	.2860	-1.2518	0	0
20-24	6.83	330622	.2066	-1.5770	1	-1.5770
25-29	7.11	316618	.2246	-1.4934	2	-2.9868
30-34	5.33	302421	.1762	-1.7361	3	-5.2083
Sum				-8.7192		-4.6891

$$\text{The growth } r = \frac{\text{Sum of Column (7)}}{5 \times 28} = \frac{-4.6891}{140} = .0335$$

Note that 5 is the scale used and 28 is the sum of squares of X values

If we carry out similar calculation for females the value of r comes out as .0295.

It can be noted that level 11 of North model used in the above exercise produces a slightly higher growth rate for males and a slightly lower growth rate for females. The higher age exaggeration for males than females has resulted in the higher value or r. Actually a level slightly higher than 11 has been estimated as applicable to the data for the period 1967-78.

To obtain the birth rate, we calculate

$$V = \frac{\text{Sum of col (5)}}{7} + 17.5\ (r) \text{ and then take anti ln}$$

For males the values come out as

$$V = -\ 1.2456 + .5362 = -.6594 \text{ for males}$$

Taking anti ln, we get the male birth rate as 51.7. Similarly, the female birth rate is 48.2 with an average near 50. With an estimated growth rate of about 3% this implies a death rate of 2% which also fits in with a level of slightly more than 11 as obtained above.

Note that in the equation for V, the value 7 and 17.5 are respectively the number of age groups used and the mid point of the transformed origin class.

8.2 ESTIMATING MORTALITY FROM CHILD SURVIVAL - BRASS METHOD

The proportions of dead children are tabulated by women grouped in five years age categories before they are converted into probabilities of dying between birth (age 0) and various childhood ages using the method illustrated in Table 21 .

Table 20 shows the number of women with the total, average and proportion of children surviving. The proportions are obtained by dividing the number surviving by the number ever born alive as given in Table 21 .

Table 2o NUMBER OF WOMEN BY FIVE YEARS OF AGES BY NUMBER, AVERAGE NUMBERS AND PROPORTION OF CHILDREN SURVIVING - Tanzania by rural/urban, 1978

No. of women	Rural	742298	607677	588069	436116	391071	311718	281506
	urban	135641	134842	115480	68682	55567	36988	30445
	Total	877939	742519	703549	504798	446638	348706	311951

.../120

Table (ctd..)

No. of children surviving	Rural	257441	1016283	1707858	1685591	1691829	1311487	1127348
	Urban	48520	212541	329972	260036	224540	142851	112466
	Total	305961	1228824	2037830	1945627	1916369	1454338	1239814
Average No. of children surviving	Rural	.347	1.672	2.904	3.865	4.326	4.303	4.005
	Urban	.358	1.576	2.857	3.786	4.041	3.862	3.694
	Total	.348	1.655	2.896	3.854	4.291	4.257	3.974
Proportion of children surviving	Rural	.8156	.8070	.7860	.7634	.7431	.7172	.6840
	Urban	.8529	.8551	.8399	.8136	.7914	.7647	.7265
	Total	.8213	.8150	.7942	.7704	.7485	.7216	.6877

Brass (1964, 1975) developed a set of multipliers for converting the proportion of dead children into estimates of the probability of dying and are given in the next page.

The estimation procedure assumes that:-

1. Age-specific fertility and mortality rates have remained constant over the required age range and time period;

2. The experience of the surviving women represents the total number of women exposed to the risk of births and deaths of children.

If we denote the proportion of children dead among those ever-born alive by women in successive five-year intervals as $D_{(i)}$ (where i ranges from 1 to 7 and represents age groups 15-19, 20-24, ---------- and 49-49 respectively), Brass multipliers ($K_{(i)}$) convert $D_{(i)}$ into estimates of $q_{(x)}$ (the probability of dying between birth and exact age x) as follows:-

$$K_{(i)}\ D_{(i)} = q_{(x)} \quad \text{-------------} \quad (1)$$

The approximate equality of $q_{(1)}$ and $D_{(1)}$, $q_{(2)}$ and $D_{(2)}$ --------------------- $q_{(20)}$ and $D_{(7)}$ is affected by variations in the age pattern of fertility among different populations more strongly than the variations in the age pattern of mortality. Thus the Brass multipliers (K_i) differ for different fertility patterns. In a population with an earlier start of childbearing the children ever-born to women aged 20-24, for example, are older than in a population where the onset of childbearing is late. To accommodate this variation in fertility patterns, Brass provided four indices for determining whether a population has early or late childbearing, and these are the ones used to select the multipliers. Even though

TABLE MULTIPLYING FACTORS FOR ESTIMATING CUMULATED FERTILITY

Exact limits of age interval								
15-20	1.120	1.310	1.615	1.950	2.305	2.640	2.925	3.170
20-25	2.555	2.690	2.780	2.840	2.890	2.925	2.960	2.985
25-30	2.925	2.960	2.985	3.010	3.035	3.055	3.075	3.095
30-35	3.055	3.075	3.095	3.120	3.140	3.165	3.190	3.215
35-40	3.165	3.190	3.215	3.245	3.285	3.325	3.375	3.435
40-45	3.325	3.375	3.435	3.510	3.610	3.740	3.915	4.150
45-50	3.640	3.895	4.150	4.395	4.630	4.840	4.985	5.000
f1/f2	.036	.113	.213	.330	.460	.605	.764	.939
m(years)	31.7	30.7	29.7	28.7	27.7	26.7	25.7	24.7

Table. Multiplying Factors to Estimate the Proportion of Children Born Alive but Dying at age X, q(x), by 5 year Age Group of mother.

15-20	q(1)	0.859	0.890	0.928	0.977	1.041	1.129	1.254	1.425
20-25	q(2)	0.938	0.959	0.983	1.010	1.043	1.082	1.129	1.188
25-30	q(3)	0.948	0.962	0.978	0.994	1.012	1.033	1.055	1.081
30-35	q(5)	0.961	0.975	0.988	1.002	1.016	1.031	1.046	1.063
35-40	q(10)	0.966	0.982	0.996	1.011	1.026	1.040	1.054	1.069
40-45	q(15)	0.938	0.955	0.971	0.988	1.004	1.021	1.037	1.052
45-50	q(20)	0.937	0.953	0.969	0.986	1.003	1.021	1.039	1.057
50-55	q(25)	0.949	0.966	0.983	1.001	1.019	1.036	1.054	1.072
55-60	q(30)	0.951	0.968	0.985	1.002	1.020	1.039	1.058	1.076
60-65	q(35)	0.949	0.965	0.982	0.999	1.016	1.034	1.052	1.070
Guide to Selection of multipler									
	P_1/P_2	0.387	0.330	0.268	0.205	0.143	0.090	0.045	0.014
	P_2/P_3	0.516	0.577	0.535	0.490	0.441	0.421	0.344	0.271
	m	24.7	25.7	26.7	27.7	28.7	29.7	30.7	31.7
	m'	24.2	25.2	26.2	27.2	28.2	29.2	30.2	31.2

technically the method produces survival probabilities upto age 35, because of (i) age reporting errors (ii) omission of children and (iii) changing patterns and levels of fertility and mortality over such a long period of time, it is advisable to accept only values of q_x upto age 5. The value of q_1 is also affected by small sample fluctuations, reporting errors and heterogeneity of the group.

The indices used for selecting the multipliers are:-

(1) The ratio of the average number of children ever-born by women aged 15-19 (P_1) to the average number of children ever-born to women aged 20-24 (P_2) or P_1/P_2;

(2) The ratio of the average number of children ever-born by women aged 20-24 (P_2) to the average number of children ever-born by women aged 25-29 (P_3) or P_2/P_3;

(3) The mean age of the fertility schedule, (m); and
(4) The median age of the fertility schedule, (m^1).

The first index (P_1/P_2) is a measure of when fertility starts and how fast it rises with age and is therefore used to select the multipliers for converting $D_{(1)}$ into $q_{(1)}$, $D_{(2)}$ into $q_{(2)}$ and $D_{(3)}$ into $q_{(3)}$. Where P_2/P_3 is used as an index it may be used for selecting the multipliers for the whole range of $D_{(i)}$. P_2/P_3 is said to be a more satisfactory index than P_1/P_2 because P_1 is sensitive to age reporting errors at the onset of childbearing and to sampling fluctuations due to small numbers of births . The inclusion of P_3 in P_2/P_3 makes the index quite satisfactory for selecting multipliers for older age groups of women, too. Another index for selecting multipliers for older age groups of women is m or m^1 if m cannot be calculated.

Table 21 illustrates the application of the Brass method derive mortality level in Tanzania.

Table 21 BRASS METHOD FOR ESTIMATING CHILD MORTALITY, Tanzania 1978

Age of women	Proportion of children dead (1)	Multiplier (2)	Probability of dying, q_x (3)=(1)x(2)	Age of child, x	Mortality level (North Model)
15-19	.1787	.908	.1623	1	9.5
20-24	.1850	.971	.1796	2	11.1
25-29	.2058	.970	.1996	3	11.3
30-34	.2296	1.016	.2333	5	11.2
35-39	.2515	1.026	.2580	10	11.6
40-44	.2784	1.004	.2795	15	11.5
45-49	.3123	1.003	.3132	20	11.0

P_i/P_2 = .199, P_2/P_3=.557 , m= 28.7

Note: The first three multipliers are obtained from Table on page 120a using the index P_2/P_3 = .557 and the last from using m.

8.3 SULLIVAN AND TRUSSEL METHODS

Sullivan (1972) examined the Brass method using regression analysis. He used observed fertility schedules, rather than a model which Brass used, together with model life tables from the Coale-Demeny system. Although he concluded that estimates obtained by using the Brass method are satisfactory, the regression equations he developed provided a simpler procedure of estimation. Further, his method is more flexible than the Brass method in that it provides equations for four different patterns of mortality which are included in the Coale-Demeny Model Life Table system. The regression equations are of the form

$$q_{(x)}/D_{(i)} = A_{(i)} + B_{(i)}(P_2/P_3)$$

Page 122 a gives the standard regression coefficients k(i).

Trussell (1975) worked out a third set of multipliers using regression equations, too. His procedure is different from that of Sullivan in that he used data generated from model fertility schedules (instead of observed schedules which Sullivan used). The methods are based on the same theory, but they yield somewhat different multipliers because their data bases are different.

The multipliers $k_{(i)}$ required to adjust the proportions dead for the effect of the age pattern of childbearing are calculated from P_1/P_2 and P_2/P_3 together with standard regression coefficients as follows:

$$k_{(i)} = A_{(i)} + B_{(i)}(P_1/P_2) + C_{(i)}(P_2/P_3)$$

The regression coefficients $A_{(i)}$, $B_{(i)}$ and $C_{(i)}$ are taken from the appropriate family of model life tables (North, South, East and West) selected as a suitable reflection of the age pattern of mortality in the observed population. These are given in page 122 b. The methods are illustrated on tables 22 and 23 respectively.

Table 22 SULLIVAN METHOD FOR ESTIMATING CHILD MORTALITY, Tanzania 1978

Age of women	Proportion of dead children, D_i (1)	Multiplier $k_{(i)}$ (2)	Probability of dying, q_x (3)=(1)x(2)	Age of child x	Mortality level from North Model
20-24	.1850	.949	.1756	2	11.3
25-29	.2058	.892	.1835	3	12.1
30-34	.2296	.916	.2103	5	12.2

Note: The multipliers are obtained from the table of Sullivan coefficients for P_2/P_3 = .557. For example, at age 25-29 the multiplier is:

$$1.17 - .5(.557) = .892$$

Table

Regression coefficients to be used in estimating adjustment factors $k(i)$ for the Sullivan variant when the data are classified by age of mother.

(1) MODEL	(2) AGE GROUP	(3) i	(4) $q^{(x)}/D(i)$	(5) $A(i)$	(6) $B(i)$
North	20 - 24	2	$q^{(2)}/D(2)$	1.30	- 0.63
	25 - 29	3	$q^{(3)}/D(3)$	1.17	- 0.50
	30 - 34	4	$q^{(5)}/D(4)$	1.15	- 0.42
South	20 - 24	2	$q^{(2)}/D(2)$	1.33	- 0.61
	25 - 29	3	$q^{(3)}/D(3)$	1.20	- 0.44
	30 - 34	4	$q^{(5)}/D(4)$	1.14	- 0.32
East	20 - 24	2	$q^{(2)}/D(2)$	1.26	- 0.44
	25 - 29	3	$q^{(3)}/D(3)$	1.14	- 0.33
	30 - 34	4	$q^{(5)}/D(4)$	1.11	- 0.26
West	20 - 24	2	$q^{(2)}/D(2)$	1.30	- 0.54
	25 - 29	3	$q^{(3)}/D(3)$	1.17	- 0.40
	30 - 34	4	$q^{(5)}/D(4)$	1.13	- 0.33

Table [illegible]

Regression coefficients to be used in estimating adjustment factors k(i) for the Trussell variant when the data are classified by age of mother.

(1) MODEL	(2) AGE GROUP	(3) i	(4) Q(X)/ D(i)	(5) A(i)	(6) B(i)	(7) C(i)
North	15 - 19	1	q(1)/ D(1)	1.1119	-2.9237	.8507
	20 - 24	2	q(2)/ D(2)	1.2390	-0.6865	- .2745
	25 - 29	3	q(3)/ D(3)	1.1884	0.0421	- .5156
	30 - 34	4	q(5)/ D(4)	1.2046	0.3037	- .5656
	35 - 39	5	q(10)/D(5)	1.2586	0.4236	- .5898
	40 - 44	6	q(15)/D(6)	1.2240	0.4222	- .5456
	45 - 49	7	q(20)/D(7)	1.1772	0.3486	- .4624
South	15 - 19	1	q(1)/ D(1)	1.0819	-3.0005	.8689
	20 - 24	2	q(2)/ D(2)	1.2846	-0.6181	- .3024
	25 - 29	3	q(3)/ D(3)	1.2223	0.0851	- .4704
	30 - 34	4	q(5)/ D(4)	1.1905	0.2631	- .4487
	35 - 39	5	q(10)/D(5)	1.1911	0.3152	- .4291
	40 - 44	6	q(15)/D(6)	1.1564	0.3017	- .3958
	45 - 49	7	q(20)/D(7)	1.1307	0.2596	- .3538
East	15 - 19	1	q(1)/ D(1)	1.1461	-2.2536	.6259
	20 - 24	2	q(2)/ D(2)	1.2231	-0.4361	- .2245
	25 - 29	3	q(3)/ D(3)	1.1593	0.0581	- .3479
	30 - 34	4	q(5)/ D(4)	1.1404	0.1991	- .3487
	35 - 39	5	q(10)/D(5)	1.1540	0.2511	- .3506
	40 - 44	6	q(15)/D(6)	1.1336	0.2556	- .3428
	45 - 49	7	q(20)/D(7)	1.1201	0.2362	- .3268
West	15 - 19	1	q(1)/ D(1)	1.1415	-2.7070	.7663
	20 - 24	2	q(2)/ D( )	1.2563	- .5381	- .2637
	25 - 29	3	q(3)/ D(3)	1.1851	.0633	- .4177
	30 - 34	4	q(5)/ D(4)	1.1720	.2341	- .4272
	35 - 39	5	q(10)/D(5)	1.1865	0.3080	- .4452
	40 - 44	6	q(15)/D(6)	1.1746	0.3314	- .4537
	45 - 49	7	q(20)/D(7)	1.1639	0.3190	- .4435

The Trussell method has been modified to yield regression coefficients for age groups 35-39, 40-44 and 45-49.

Table 23 TRUSSEL METHOD FOR ESTIMATING CHILD MORTALITY, Tanzania 1978

Age of women	Proportion of dead children, D_i (1)	Multiplier $k_{(i)}$ (2)	Probability of dying, q_x (3)=(1)x(2)	Age of child, x	Mortality level from North Model
20-24	.1850	.949	.1756	2	11.3
25-29	.2058	.910	.1873	3	11.9
30-34	.2296	.950	.2181	5	11.9

$P_1/P_2 = .199$, $P_2/P_3 = .557$

Note: The multipliers are obtained from the table of Trussel coefficients for $P_1/P_2 = .199$ and $P_2/P_3 = .557$.

For example, at age 25-29, the multiplier is:

$$1.1884 + .0421(.199) - .5156(.557) = .910.$$

These methods of estimation assume that fertility and mortality have been constant in the past. Therefore some biases can occur if fertility and mortality have been changing in the past. These biases are, however, very minor compared to errors created by misreporting. One other factor that may affect the results is the assumption that the risk of a child dying is a function only of the age of the child and not other factors such as mother's age and birth order of the child. Again biases caused by this factor are considered to be generally minor.

It can be noticed that the levels indicated by q_2, q_3 and q_5 from all of the methods fluctuate. This is due to age errors, reporting biases with regard to child survival. To obtain a smoothed value, we can take an average of these levels. For example, for the Sullivan method the average level = (11.3 + 12.1 + 12.2)/3 = 11.9. Another method proposed by Brass is the logit smoothing of l_x values using an appropriate standard. We have illustrated the smoothing of l_x values based on all the 3 methods in Table 24. The standard used is Coale-Demeny North model. It can be seen that the levels indicated by the smoothed values are same for each method but differ from one method to another. Probably an average of the 3 smoothed values i.e. (11.2 + 11.7 + 11.7)/3 = 11.5 is near the true level of mortality in the country in the period between the two censuses, i.e. around 1973.

Table 24 LOGIT SMOOTHING OF BRASS, SULLIVAN AND TRUSSEL ESTIMATES FOR TANZANIA - USING NORTH 12 AS STANDARD

x	1x	Y_x	Y^S_x	Brass Difference $Y_x - Y^S_x$	Adjusted Y_x	Adjusted l_x	Level from north model
2	.8204	-.7595	-.8161	.0566	-.7649	.8220	11.2
3	.8004	-.6944	-.7389	.0445	-.6877	.7982	11.2
5	.7667	-.5949	-.6471	.0525	-.5952	.7668	11.2

Table (ctd..)

			Sullivan				
2	.8244	-.7732	-.8161	.0429	-.7946	.8305	11.7
3	.8165	-.7464	-.7389	.0075	-.7174	.8076	11.7
5	.7897	-.6616	-.6474	.0142	-.6259	.7776	11.7
			Trussel				
2	.8244	-.7732	-.8161	.0429	-.7971	.8312	11.7
3	.8127	-.7338	-.7389	.0051	-.7199	.8084	11.7
5	.7819	-.6384	-.6474	.0090	-.6284	.7785	11.7

Note: Y_x^s values are logits of 1_x values from North model level 12.

9. ANALYSIS OF FERTILITY DATA

Introduction

Even though the situation regarding fertility information may be similar to that of mortality, surveys have been successful in obtaining levels and sometimes patterns of fertility. Direct information through registration system is still inadequate and hence we have to base our estimates on indirect measures.

Two such techniques based on age structure and report of current and retrospective fertility will be used.

9.1 AGE STRUCTURE - COALE-DEMENY METHOD

As noted earlier, when fertility and mortality could be assumed to have remained constant for a very long time, the age distribution of a population tends to stability. This is approximately true also when fertility is more or less constant but mortality may be falling but slowly.

Thus an age distribution can be compared with stable model age distributions to estimate the inherent parameters. Since age-sex reporting errors may have vitiated a reported age distribution, it has been suggested by Coale and Demeny that it is better to compare cumulated age distributions rather than the actual reported. For removing some of the biases broughtin by age reporting errors, it is recommended to use the first 7 or 9 age groups for comparison. Corresponding to the observed cumulative values (percentages) the method consists of matching with model stable populations with the given growth rate, r and GRR. These would result in a series of 7 or 9 mortality levels corresponding to r and GRR. The average of the medians based on r and GRR is taken as the estimated level of mortality of the population. From the stable population, the other parameters like birth rate could easily be read off.

The method also can be based on an estimated level of mortality (as depicted by 1_2 or by smoothed 1_x values by Brass or other methods).

Thus the method produces estimates of fertility and mortality for given r or GRR or produces estimates of fertility and growth rate for given mortality level.

The method is illustrated in Table 25 using the estimated growth rate of 3.1% and GRR of 3.5 basing the comparison on North model stable populations.

Table 25 COALE-DEMENY METHOD FOR ESTIMATING VITAL PARAMETERS FROM AGE, DATA, Tanzania 1978

Age upto	Cumulated %		Level from North model with r= 3.1%		Level from North model with G.R.R.= 3.5	
	M	F	M	F	M	F
5	18.14	18.16	16.2	14.9	4.5	6.5
10	34.34	33.98	13.1	12.6	10.0	10.8
15	46.76	45.57	13.5	14.1	9.8	9.3
20	56.56	55.41	14.8	15.1	8.5	8.4
25	63.39	63.73	18.3	15.8	5.4	7.8
30	70.50	71.61	18.7	14.7	4.7	8.5
35	75.83	77.27	20.4	15.5	3.0	7.5
40	80.95	82.27	20.5	15.6	2.0	3.0
45	84.69	86.18	22.0	16.4	1.0	5.7
Median level			18.3	15.1	4.7	7.8

The median levels are too high based on r and too low based on GRR. An average of these two (level 11.5) seems near the true situation.

Corresponding to level 11.5 and r = 3.1% the birth and death rates from stable models are: 51.0 and 20.0. Using GRR = 3.5 the birth and death rates come out as: 50.4 and 19.9.

9.2 REVERSE SURVIVAL RATIO METHOD

The idea of this method is based on the premise that the reported population are the survivors of births of past periods. Hence if we survive a population for the 5 years previous to a census by using appropriate life tables, then the children aged 0-4 at census time would result in births of the 0-4 year period previous to census. Similarly, if we survive these again for a further 5 years, those aged 5-9 years at census time would result in births of the period 5-9 years before census. These births and populations could then be utilized to estimate birth, death and growth rates for the decade previous to census.

Table 26 shows the results of the calculation done for the enumerated female population of 1978 in Tanzania. It can be noticed that for carrying out the procedure we need life tables for the two periods. In the absence of an available life table, we have used North model life tables with levels 12 for 1973-78 and 11 for 1968-73, i.e. level 11.5 during 1968-78. A slightly higher or lower level will have only marginal effect on the results. We also have used only the percentage age distribution instead of the actual numbers. For our purpose the choice is immaterial but the smaller values of percentages makes calculations easy.

Table 26 REVERSE SURVIVAL RATIO METHOD - FEMALE POPULATION, Tanzania 1978

Population 1968 (1)	Survival ratio 1968-73 (2)	Population 1973 (3)	Survival ratio 1973-78 (4)	Population % 1978 (5)	Age
20.69		21.39			Birth
13.20	.833	17.23	.849	18.16	0-4
10.47	.908	11.99	.918	15.82	5-9
8.79	.963	10.08	.967	11.59	10-14
8.38	.973	8.55	.976	9.84	15-19
6.08	.970	8.13	.973	8.32	20-24
5.43	.966	5.87	.969	7.88	25-29
[illegible]	.960	5.21	.96[illegible]	5.66	30-3[illegible]
3.89	.954	4.10	.959	5.00	35-39
3.00	.949	3.69	.953	3.91	40-44
2.28	.944	2.83	.948	3.50	45-49
2.[illegible]	.93[illegible]	2.13	.93[illegible]	2.66	50-54
1.76	.915	2.19	.921	1.96	55-59
[illegible]	.882	1.55	.890	1.91	60-64
[illegible]	.678	3.95	.839	1.30	65-69
[illegible]	.719	[illegible]	.617	2.[illegible]	70 +
75.7[illegible]	—	97.[illegible]	—	[illegible]	Total

Notes (a) The survival ratios for 1968-73 and 1973-78 are respectively from North model level 11 and 12.

(b) Column (3) is obtained by dividing column (5) by column (4).
(c) Column (1) is obtained by dividing column (3) by column (2).
(d) The sums in the last row exclude the births.

The birth rates for 1973-78 and 1968-73 are obtained by dividing the births by the average population and dividing by 5 (the no. of years) e.g. the birth rate during 1973-78 is:

$$B.R_1 = \frac{2(21.39)}{5(87.49+99.98)} = .0456 \text{ or } 45.6 \text{ per } 1000$$

Similarly birth rate during 1968-73 is:

$$B.R_2 = \frac{2(20.69)}{5(75.79+87.49)} = .0507 \text{ or } 50.7 \text{ per } 1000$$

The growth rates can be obtained by the exponential growth formula, e.g. growth rate between 1973-78 is:

$$r_1 = \frac{1}{5} \ln \frac{99.98}{87.49} = .0267$$

and growth rate between 1968-73 is:

$$r_2 = \frac{1}{5} \ln \frac{87.49}{75.79} = .0287$$

The death rates are the difference between birth and growth rates, e.g.

$$d_1 = .0456-.0267 = .0189 \text{ or } 18.9 \text{ per } 1000$$

and $\quad d_2 = .0507-.0287 = .0220 \text{ or } 22.0 \text{ per } 1000$

The exaggeration of ages can be noted from the huge expected populations at old ages. This exaggeration has resulted in larger population figures in the past and consequently has reduced the growth rate. If age reporting had been better, then the birth rates would be higher and growth rates also higher but death rates lower.

9.3 BRASS P/F RATIO METHOD

In Tanzania, from past censuses and surveys, two types of fertility data were collected. The two types of data both recorded by age of mother are births which occured during the year or date of most recent birth (i.e. current period) and the number of children ever born to each woman (i.e. life-time).

The information on current births is obtained by questioning the women about whether they have borne children during the twelve months before the census. Age specific fertility rates can be obtained by dividing the number of births to mothers by the corresponding total women in the population for each age group. Measures based on such specific rates will be referred to as "current" and the in dices of fertility obtained from the mean number of children ever bor n to women of each age will be called "retrospective".

The two types of measure can be used to detect and allow for errors in the data because of the logical relationship between them.

As a cohort of women moves through life, the mean number of children ever born at each exact age is equal to the cumulative total of age-specific fertility rates to that age, if we assume that the women dying have the same fertility as those surviving. In other words, the availability of information about both retrospective fertility and current fertility can, therefore, be seen to make possible a powerful consistency check, whereby current fertility rates can be accumulated and compared with the average children ever born. This comparison of retrospective with current fertility data can also provide a method of adjustment for cases where the data are distorted by typical errors. It has already been said that information on children ever born is frequently distorted by omission, but that such omission, perhaps of children who died long ago or of those who have left the parental home, is most marked for older women; the reports of younger women up to say, age 30 or 35 may be fairly reliable.

Information on current fertility from a question on births in the year before the census may be distorted by a misperception of the duration of the reference period so that reported births correspond to an ill defined period whose average length may be longer or shorter than one year. These errors in the information on current fertility may be assumed to be roughly constant by age, particularly in the case of the reference period error, and the age pattern of fertility rates is accepted though their level is distorted. This implies that we actually have two types of information: one on the level of fertility - from reports of children ever born of younger women - that may be approximately correct, and information on the age pattern of fertility - from reports of women about current fertility - that may be relatively free of distortion. Cumulated current fertility may be compared with reported retrospective fertility for women up to age 30 or 35 in order to obtain an adjustment factor for the level of the current fertility rates, which, once adjusted for the level, provide a best estimate of actual current fertility.

The essence of the Brass fertility estimation procedure is, therefore, the adjustment of the level of fertility using the information on age pattern of fertility derived from information on recent births and the level of fertility indicated by the average number of children ever born by women in the age groups 20 to 24, 25 to 29 and perhaps 30 to 34.

Data required for the Brass P/F Ratio Method are as follows:

(1) the number of children ever born by age of mother; (2) the number of children born during the year preceding the census and classified by age of mother; (3) the total number of women in each age group; and (4) the total population if the birth rate is to be estimated.

The method is illustrated in Table 27 for Tanzania, 1978, using data from Table 16.

Table 27 BRASS P/F RATIO METHOD, Tanzania 1978

Age of women, x (1)	Age specific fertility rate, f_x (2)	G_i (3)	Multiplier (4)	F_i (5)	Average Parity, P_i (6)	P/F (7)
15-19	.135	-	2.241	.303	.424	1.40
20-24	.309	.675	2.881	1.565	2.031	1.30
25-29	.297	2.220	3.030	3.120	3.647	1.17
30-34	.246	3.705	3.120	4.472	5.003	1.12
35-39	.180	4.935	3.245	5.519	5.733	1.04
40-44	.088	5.835	3.510	6.144	5.899	.96
45-49	.037	6.275	4.395	6.438	5.779	.90

$f_1/f_2 = .437$ m = 28.7

Note: (a) Column (3) i.e. G_i values are obtained from column (2) by first multiplying the f_x values by 5 and successively cumulating, e.g.,

$$3.705 = 5(.135) + 5(.309) + 5(.297)$$

(b) The multipliers in column (4) are obtained from table on page 120 a using f_1/f_2 for the first three and m for the last four values

(c) Column (5) i.e. F_i values are obtained by multiplying values in column (2) by values in column (4) and adding the value in column (3), e.g.,

$$4.472 = .246(3.120) + 3.705$$

(d) Column (7) is the ratio of columns (6) and (5)

To adjust the level of fertility usually the factor P_2/F_2 is used but when age errors are predominant as in Tanzania, it seems advisable to use an average of P/F values for several ages. Even this method does not seem to be satisfactory because if we use ages 20-39 the correction factor comes out as 1.16 i.e. the reported current fertility should be increased by 16% resulting in a TFR of 7.5 and CBR of 53.2. Apparently this looks too high and perhaps the method has not produced reasonable results.

Two formulas based on parity values used to estimate TFR due to Coale-Demeny and Brass respectively seem to work better. The Coale-Demeny formula is:

$$TFR = P_3^2/P_2 = \frac{(3.647)^2}{2.031} = 6.5$$

and the Brass formula is:

$$TFR = P_2(P_4/P_3)^4 = 2.031(5.003/3.647)^4 = 7.2$$

The actual value of TFR may be between these 2 values. A round figure of 7 is recommended corresponding to a CBR of 50.

10. POPULATION DISTRIBUTION, MIGRATION AND URBANIZATION

10.1 POPULATION DISTRIBUTION

Introduction

The population of a country is both a productive resource and the resource user. It demands essential services such as education, health, food, clean water supply, transport, housing, law and order, and participation in the nation's affairs. Because of these needs, it is neceesary for a government to know how many people there are and where they are spatially located. The way people are dispersed or concentrated in the national space affects decisions on the kinds of development programmes and where to locate them.

Only some of the most common concepts, methods and techniques used in the analysis of the distribution and density of population will be highlited. These concepts include relative distribution of population, measures of population dispersion and concentration, and average density of population.

Population Distribution

Population distribution refers to the way people are spread within the area available to them for exploitation and settlement. To analyse distribution, one needs data on both absolute numbers of people and a well defined territory.

Several methods of showing population distribution have been developed. The most visually effective technique of showing how a population is spread over its national area is by means of a dot map-on which dots are used to represent a given number of people. The 1967 population distribution of Tanzania was shown on a dot map in the Atlas of Tanzania (1976, p.16).

The total population of Tanzania in 1978 was 17,512,610 persons occupying an area of 885,987 square kilometres. Out of these totals, Zanzibar and Pemba had 476,111 persons on 2,644 square kilometres. Comparable figures for 1967 were 12,313,496 people for the whole of Tanzania and 354,815 persons for the Islands.

Two aspects of the distribution of population in Tanzania are important. First, the distribution between rural and urban areas shows that in 1978 there were 2,412,900 people or 13.8 percent of the total population living in towns. Out of this urban population, 2,257,121 were on the mainland and 154,979 in Zanzibar and Pemba representing 13.3 percent and 29.4 percent of their populations respectively. In 1967, 6 percent of the total population lived in towns with Mainland Tanzania having only 5.4 percent and in the Islands the figure was 27.4. These figures show a very rapid rate of urbanization especially for Mainland Tanzania.

The second aspect of the distribution of population in Tanzania is the uneven distribution in the rural areas. The basic pattern is that of heavy concentrations of population in the highland and coastal areas contrasting with very sparse distribution in the central and plateau areas. This pattern has persisted over the past 50 years with slow shifts to the low-density areas. The reasons for the persistence of this basic uneven pattern are mainly environmental and include the total annual rainfall and its reliability and variability; the presence of the tsetse fly over vast areas of the country; predominance of infertile and leached soils; and the low level of technology.

The unevenness of the distribution of population is highlighted when the distribution is analyzed at regional and district level. In table 12, regions have been arranged in order of density from highest to lowest and the cumulative percentage contributions to the total population and area of Tanzania have been computed. It will be noted that at regional level, 64 percent of the total population occupied 36 percent of the total area. This aspect of population concentration becomes more obvious as the unit of analysis becomes smaller. For example, at the district level, 65 percent of the population occupied 28 percent of the area, and at the ward level 65 percent of the population occupied 20.6 percent of the total area. Compared to the 1967 analysis, there has been a slight tendency towards less concentration or there has been more spreading out of the population in Tanzania as the comparable figures at district level were 63 percent of the population occupying 25 percent of the area and 65 percent occupying 18.8 percent of the area at local (ward) level.

The degree of concentration of a population depicted visually on Lorenz curves can be seen for data on Table 12 shown in Fig. 10. At the regional level the index of concentration was .39 which is not very high.

The relative distribution of population or the spatial patterns of distribution can be noted from the index of distribution of population obtained by dividing the population density of the district (or any other sub-national unit) by the national average density. This index shows the contribution of population of each district relative to its area. When the index is 1.0 the proportion of the total population living in that district is equal to the proportion of the total area occupied by that district. Values above or below 1.0 indicate respectively that the proportion of the population exceeds or falls short of the proportion of the total area.

Density of Population

The concept of population density, relating numbers of people to the space occupied by them, is a useful one in making comparisons between areas and over different time periods. Crude or average density usually refers to the number of people per administrative unit,

and the smaller the unit, the more meaningful the average figure obtained. Sometimes, it may be necessary to calculate the density for the inhabited areas only and sometimes the density of the agricultural population only. The agricultural density of population provides more interesting comparisons over small areas and between countries.

The density of population for Tanzania in 1978 was 19.8 persons per square kilometre and 193.5 for Zanzibar and Pemba. The density for the Mainland alone was 1[illegible].7 persons per square kilometre. Comparable figures in 1967 were 13.5 and 134.2 per square kilometre for the Mainland and the Islands respectively. There are great variations in the density of rural population in Tanzania as can be seen from Fig.9 In the zones of population concentration, rural densities per district vary from 300 persons per square kilometre in Chakechake and 233 per square kilometre in Moshi rural, to 135 in Musoma, 96 in Rungwe and 46 in Shinyanga districts.

Parts of central Tanzania and parts of the Southern Highlands had moderate densities ranging from 30 per square kilometre in Masasi and 26 in Dodoma, to 23 in Iringa and 17 per square kilometre in Sumbawanga. More than half of the national area (56 percent) had very low densities of less than 15 persons per square kilometre and included the miombo woodlands and bushlands of the central plateau, south-eastern Tanzania and Masailand. Most of these areas do not only suffer from unreliable rainfall and leached soils but are also tsetse infested and contain most of the national parks and game reserves.

One of the reasons for the uneven population distributio n is migration. It is also pos sible that areas difficult for human habitation may have low er natural growth. We shall now consider estimation of migration.

M I G R A T I O N

Basic definitions:

Migration is one of the basic demographic variables. However because of its complexity it is not easy to define. A generally accepted definition is that, migration is the movement of population involving the change of place of abode or place of usual residence and the crossing of a defined boundary. Migration is an event that occurs in time and hence the temporal aspect is also important as is the spatial. When such movements occur within a country they are referred to as internal migration, while if they involve crossing national boundaries they are referred to as international migration. Emigration and immigration refer to migration away from and into a country while out-migration and in-migration refer to internal migration from or to a given area within a country. All moves or migrations with a common area of origin and destination are referred to as migration streams.

Measurement of migration

Migration can be measured either directly or indirectly. Direct measurement requires a system of recording movements as they occur. The most complete migration statistics are developed from

population registers in which all changes of residence are recorded. However for most countries, cenus data have been the major source of information on internal migration. For international migration, statistics collected at border posts is a source but in countries with large borders and with several border countries, this may not be accurate.

The census data on internal migration are obtained directly by including a question on migration. The usual direct questions seek information on: place of birth, place of last residence, duration of residence in the place of enumeration and place of residence on a specific date before the census. Using such information the population in an area can be classified into two groups: migrants and non-migrants. The direct questions bearing on migration included in the 1978 census were (1) place of birth (2) place of residence in 1977 (3) place of residence in 1967.

Indirectly, migration may be estimated by comparing of total population counts for component areas in two censuses. The commonly used methods are: (1) the vital statistics method which consists of computing the difference between total population change, as assessed from two censuses and natural increase during the intercensal period; and (2) the survival Ratio Method which is used to estimate net migration by age; either derived from suitable life tables or from the comparison of successive censuses, say the 1967 and 1978 censuses.

Methods of Indirect Estimation

The indirect estimation of migration is based on the knowledge that population increment between any two dates for any given geographic area is the result of natural increase (births minus deaths) and net migration.

Given the population of an area at two points in time and an estimate of natural increase during the interval, we can calculate the number that would be expected at the end of the interval in the absence of migration. The difference between the observed and the expected change, gives an estimate of net change due to migration.

10.2 NATIONAL GROWTH RATE METHOD

This is a rough but simple and commonly used method of estimating net migration in an area. The idea is based on the comparison of the regional growth to the national growth and any difference is attributed to migration. It assumes that the natural growth rates in the regions are same as that in the country and that enumeration completeness is similar between regions.

The method consists in calculating regional and national growths for the intercensal interval by the simple formula:

$r = (P_n - P_o)/P_o$; where P_o and P_n denote respectively the population at the earlier and later censuses.

.../134

The migration rate for region i is calculated as

$$m_i = (P_n^i - P_o^i)/P_o^i - (P_n - P_o)/P_o$$

Applying this migration rate to the base population of region i we get the net migration.

Applying this method to estimate net migration in Zanzibar North, we have the population figures in Zanzibar North for males in 1967 and 1978 respectively as:

	Zanzibar	Zanzibar North
1967	177905	27974
1978	236592	38270

Hence

$$m = (38270 - 27974)/27974 - (236592 - 177905)/177905$$

$$= .368 - .330 = .038$$

Hence net migration to Zanzibar North is .038(27974) = 1063.

10.3 CENSUS SURVIVAL RATIO METHOD

If vital statistics are available and acceptable, then from the balancing equation:

$P_n = P_o + (B - D) + (I - E)$, where B, D, I and E denote respectively the births, deaths in and out-migration during the intercensal interval and P_o and P_n are the populations at the earlier and later counts, we can obtain the net migration component as:

$$I - E = P_n - P_o - B + D$$

However in countries where vital statistics are poor, as in Tanzania, this method cannot be attempted.

The growth rate method given earlier is some kind of an approximation to the vital statistics method.

Yet another adaptation of the balancing equation method is the census survival ratio method which we shall illustrate by using Zanzibar data. The method requires comparable age-sex distribution at the two enumerations.

Operationally, a census survival ratio is the ratio of the population aged $x + n$ at a given census to the population aged x at the census n years earlier. The ratio is then multiplied by the population aged x in a given area at the earlier count and the expected survivors are subtracted from the corresponding population enumerated at the second census to yield estimates of net migration.

Symbolically, if $P_{i, x, t}$ refers to the population in the ith region in a particular age group x at the first census (time t),

$P_{i,x+n,\ t+n}$ the corresponding population n years older (i.e. x+n) at the next census after n years and $P_{x,t}$ and $P_{x+n,\ t+n}$ refer to the corresponding population of the country as a whole, the estimate of net migration is given by the formular:

$$m_i(x) = P_{i,x+n,\ t+n} - \frac{P_{x+n,\ t+n}}{P_{x,t}} \cdot P_{i,x,t}$$

The ratio $P_{x+n,\ t+n}/P_{x,t}$ is the census survival ratio at age x.

Normally population data are compiled by five-year age groups and the intercensal interval is either 5 or 10 years. However in the case of the last two censuses in Tanzania the interval is 11 years (1967 - 1978). In this case adjustment procedures are required. These will not be discussed here. The tabulations that follow illustrate the procedure used to estimate net migration by age using census survival ratios. The age groups for 1978 have been compiled from single years in order to obtain age groups corresponding to those of the 11 year census interval. i.e. population aged 0-4 in 1967 will be aged 11-15 in 1978. Table 28 gives the intercensal survival ratios for males in Zanzibar. Table29shows the procedure for the calculation of net migration.

Table 28 MALE POPULATION - ZANZIBAR BY AGE 1967 AND 1978 AND CENSUS SURVIVAL RATIO 1967-78

AGE GROUPS 1967	MALE POPULATION	AGE GROUPS 1978	MALE POPULATION	(CSR) CENSUS SURVIVAL RATIOS, 1976-1978
0-4	32116	11-15	27372	.8523
5-9	29863	16-20	21449	.7182
10-14	14441	21-25	15303	1.0597
15-19	11246	26-30	15501	1.3784
20-24	9957	31-35	9608	.9649
25-29	13649	36-40	11931	.8741
30-34	12726	41-45	7620	.5988
35-39	10350	46-50	8457	.8171
40-44	9459	51-55	4689	.4957
45-49	6593	56-60	6974	1.0578
50-54	7621	61-65	3466	.4548
55-59	3593	66-70	4390	1.2218
60-64	6685	71-75	1830	.2737
65-69	2393	76-80	2108	.8809
70+	7213	81+	1749	.2425
All Ages	177905	Total 11+	142448	.8007

The CSR method assumes that the population is closed to international migration and the coverage at the two censuses are similar. It also assumes that regional differentials in mortality are nil. However, errors in age reporting and net immigration from abroad,

may yield a survival rate that exceeds unity for some age groups. This is an impossible value but as far as the method is concerned it is used for the purpose of estimating net migration. T he method can estimate migration only among those aged 11 years and above. Estimates of migration among children aged 0-10 years can be approximated by using relation between women and children (child women ratios). Whereas the growth rate method indicates a net in migration, the census survival ratio method shows a net out migration. Perhaps the assumptions underlying the methods are not satisfied. Usually it is advisable to check the situation with any direct evidences one may have. We shall do this by using the informati on on place of birth and place of residence in 1967 and 1977.

Table 29 ESTIMATING NET MIGRATION IN ZANZIBAR NORTH FOR MALES, 1967-78

Age in 1967	Pop. 1967 (1)	Age in 1978	C.S.R. (2)	Expected Pop. (1)x(2)	Actual Pop. (3)	Net Migration (3)-(4)
0-4	5362	11-15	.8523	4570	3957	-613
5-9	5156	16-20	.7182	3703	3426	-277
10-14	2292	21-25	1.0597	2429	2335	- 94
15-19	1556	26-30	1.3784	2145	2319	174
20-24	1296	31-35	.9649	1251	1470	219
25-29	1840	36-40	.8741	1608	1788	180
30-34	1990	41-45	.5988	1192	1173	- 19
35-39	1435	46-50	.8171	1173	1389	216
40-44	1578	51-55	.4957	782	719	- 63
45-49	1043	56-60	1.0578	1103	1219	116
50-54	1222	61-65	.4548	556	640	84
55-59	562	66-70	1.2218	687	758	71
60-64	1126	71-75	.2737	1376	347	-1029
65-69	370	76-80	.8809	326	424	98
70+	1146	81+	.2425	278	375	97

DIRECT METHODS

10.4 ESTIMATING MIGRATION FROM PLACE OF BIRTH AND PLACE OF ENUMERATION

Using data on the place-of-birth and place of enumeration it is possible to classify the population into two groups:

(a) Migrants - defined as persons who were enumerated in a place different from the place where they were born.

(b) Non-migrants - persons who were numerated in the place where they were born.

In the 1978 census tabulations, birth-places were classified by the broad categories: Same Region, Other Mainlands/or Island Region; and outside Tanzania. Birth-places were also defined by region. The compilation of birth places by place of enumeration for

Zanzibar are illustrated in Table 30 . It can be noted that 18% were migrants within Zanzibar-Pemba, 4% from mainland and about 1% from outside Tanzania. As expected, internal migrants were more at young adult ages with only few children and old persons. Migrants from mainland had less proportion of children but had sizeable older persons. The streams of life-time migration are shown in Table 31 which gives for Zanzibar the number of in-and out-migrants, the volume of net migration, the origin and destination of each stream of migration to and from each region in Zanzibar. These values are obtained from Table 32showing the population by place of birth and place of residence in 1977.

Table 30 POPULATION BY AGE AND PLACE OF BIRTH, Zanzibar and Pemba

Age	Total	PLACE OF BIRTH: Same Region	Other Island Region	Mainland Tanzania	Outside Tanzania
	476,111	368,416	86,302	17,587	3,696
0-4	90,635	79,955	9,450	1,166	64
5-9	86,431	73,826	11,188	1,357	59
10-14	54,974	45,487	8,575	835	71
15-19	46,904	34,750	10,657	1,409	81
20-24	33,172	22,067	9,355	1,658	86
25-29	30,639	19,829	9,015	1,732	157
30-34	24,446	16,659	6,289	1,218	281
35-39	19,196	12,772	4,855	1,192	363
40-44	19,751	13,663	4,509	1,172	408
45-49	13,541	8,966	2,996	1,298	352
50-54	15,114	10,675	2,959	1,115	364
55-59	7,406	5,914	1,360	745	287
60-64	12,601	9,102	2,130	903	366
65-69	5,226	3,536	947	499	244
70-74	7,122	5,339	1,041	541	201
75+	[illegible],942	6,787	1,076	765	306

Table 31 LIFE-TIME INTERNAL MIGRATION IN ZANZIBAR

Region (1)	In-Migration (2)	Out-Migration (3)	Net Migration (2) - (3)
Zanzibar North	5381	18519	-13138
Zanzibar South	6015	22131	-16116
Zanzibar West	60475	6399	54076

Table (ctd...)

Region (1)	In-Migration (2)	Out-Migration (3)	Net Migration (2) - (3)
Pemba North	6289	15772	-9483
Pemba South	5348	20687	-15339
Total	83508	83508	0

Note: The figures in column (2) are obtained by differencing the column totals and corresponding diagonal values , e.g., 6015 = 44040 - 38025. The figures in column (3) are difference between row totals and diagonal values, e.g. 6399 = 69503 - 63104.

Table 32 PLACE OF BIRTH BY PLACE OF RESIDENCE 1977, ZANZIBAR

Place of Birth	PLACE OF RESIDENCE 1977		ISLES			
	Zanzibar North	Zanzibar South	Zanzibar West	Pemba North	Pemba South	Total
Zanzibar North	67682	1221	16060	613	625	86201
Zanzibar South	607	38025	20750	382	392	60156
Zanzibar West	1512	3195	63104	920	772	69503
Pemba North	1119	773	10321	93575	3559	109347
Pemba South	2143	826	13344	4374	87938	108625
Total	73063	44040	123579	99864	93286	433832

10.5 MIGRATION ESTIMATES FROM PLACE OF PREVIOUS RESIDENCE

A question on place of residence on a fixed past date (1967 and 1977) was asked in the 1978 census. This provides a specific migration interval in which migration status is determined by a comparison of residence at two definite points in time. A migrant is defined as a person whos e residence at the census date differs from his residence at the specified prior date. Such a procedure gives a count of surviving migrants for a fixed period of time. Table 33 shows a cross-classification of migrants by place of residence in 1967 and place of residence in 1977 for Zanzibar-Pemba.

Table 34 shows the in, out and net migration streams within the islands.

We note from Tables 31 and 34 that Zanzibar West region is a gainer of migrants and all other regions are losing population. This is due to the relative better development in that region. On the other hand, Pemba South is a heavy loser perhaps due to lack of economic opportunities, social facilities and other reasons.

Table 33 ZANZIBAR INTERREGIONAL MIGRATION 1967-1977

Region of residence 1967	Region of residence in 1977					
	Zanzibar North	Zanzibar South	Zanzibar West	Pemba North	Pemba South	Total
Zanzibar North	41971	356	4892	245	249	47713
Zanzibar South	244	26443	7571	150	210	34618
Zanzibar West	755	1362	63752	786	851	67506
Pemba North	497	320	5187	57509	1324	64837
Pemba South	905	408	6385	1399	53328	62425
Total	44372	28889	87787	60089	55962	277099

Table 34 IN AND OUT MIGRATION IN ZANZIBAR, 1967-1977

Region (1)	In Migration (2)	Out-Migration (3)	Net Migration (2) - (3)
Zanzibar North	2401	5742	-3341
Zanzibar South	2446	8175	-6729
Zanzibar West	24035	3754	20281
Pemba North	2580	7328	-4748
Pemba South	2634	9097	-6463
Total	34096	34096	0

Note: Figures in column (2) are the differences between column totals and corresponding diagonal values, e.g. 24035 = 87787 - 63752. Figures in column (3) are the differences between row totals and corresponding diagonal values, e.g. 7328 = 64837 - 57509.

.../140

10.6 URBANISATION

The 1978 census definition of an urban area included all regional and district headquarters with criteria of size, class and density also being used wherever necessary. Land use criteria and nature of economic activity were other additional tests for considering areas as urban. Thus there was difference in definition between the last 2 censuses.

According to the 1978 census, 13.3% of the population of the country was urban and the urban growth rate was 10.8% per year. For Zanzibar the corresponding values were 32.6% and 3.9% and hence it can be seen that Zanzibar is comparatively highly urbanised but that the Mainland is more rapidly urbanising. Whereas in 1967 there was only one urban locality with 100000 or more population, in 1978 there were 3 out of which one had population over 500000.

There was tremendous variation in the proportion urban within the country with Dar Es Salaam having 91.3% urban as against West Lake region having only 3.4% of its population as urban.

The rather high urban growth has been the result of rural to urban migration. Table 35 and 36 show the intra-and inter-regional migration statistics for Zanzibar from the 1978 census. The volume of in, out and net migrations are shown in Table 37 from which it is obvious that rural to urban migration is a major factor for urban growth and that inter-regional migration plays an important role.

Table 35 INTRA-REGIONAL RURAL-URBAN MIGRATION IN ZANZIBAR

Place of Residence 1967	Place of Residence in 1977 Regional HQ	District HQ	Rural Area HQ
Regional HQ	55223	108	202
District HQ	211	8550	93
Rural Area	2656	1106	1299

Table 36 INTER-REGIONAL RURAL-URBAN MIGRATION

Place of Residence 1967	Place of Residence in 1977 Regional HQ	District HQ	Rural Areas
Regional HQ	5266	411	576
District HQ	1547	501	43
Rural Areas	11467	304	479

Table 37 VOLUME OF IN, OUT AND NET INTRA-REGIONAL MIGRATION, ZANZIBAR

	In migration	Out migration	Net migration
		Intra-regional	
Regional HQ	2867	310	2557
District HQ	1214	304	910
Rural Areas	295	3762	-3467
		Inter-regional	
Regional HQ	13014	987	12027
District HQ	715	1590	-875
Rural Areas	619	11771	-11152

11. POPULATION GROWTH AND STRUCTURE

11.1 POPULATION GROWTH

Introduction

Analysis of population growth require periodic and systematic information on population totals and the regular collection of birth and death records, together with data on migratory movements. Unfortunately for Tanzania relevant information in this detail is not available.

There are however ways of computing fairly accurate separate estimates for births and deaths. Crude birth rates can be computed from the child woman rates, from 2 census age distributions and from adjustments to reported current and retrospective fertility and crude death rates from census survival ratios, child mortality estimates etc. In addition, estimate of migration rate can be obtained from questions on migration or from indirect methods based on age distributions.

.../1 2

Table below shows the intercensal rate of population change and annual growth rate.

Table 38 INTERCENSAL POPULATION GROWTH

Area	Population 1967	Population 1978	Percentage change 1967/78	Annual growth rate (%) 1967/78
Mainland	11,958,654	17,036,499	42.5	3.2
Zanzibar	354,815	476,111	34.2	2.7
Zanzibar Island	190,494	270,807	42.2	3.2
Pemba Island	164,321	205,304	24.9	2.0
Tanzania	12,313,469	17,512,610	42.2	3.2

REGIONAL POPULATION GROWTH

Within the country, differential rates of population growth exist by region (see table 39). Possible factors for differential growth are differences in rates of natural increase (Birth rate - Death rate); varying intensity of internal migration, in the form of resettlement schemes or individual migration from economically unattractive areas to economically attractive ones; and differences in reception of international migrants, especially refugees.

Table 39 REGIONAL POPULATION GROWTH: MAINLAND

Region	Population 1967	Population 1978	Annual growth rate 1967/78 (%)
Dar es Salaam	356,286	843,090	7.8
Rukwa	276,091	451,897	4.5
Tabora	502,068	817,907	4.4
W. Lake (Kagera)	658,712	1,009,767	3.9
Arusha	610,474	926,223	3.8
Shinyanga	899,468	1,323,535	3.5
Mbeya	753,765	1,079,864	3.3
Ruvuma	395,447	561,575	3.2
Dodoma	709,380	972,005	2.9
Kilimanjaro	652,722	902,437	2.9
Morogoro	682,700	939,264	2.9
Kigoma	473,443	648,941	2.9
Mwanza	1,055,883	1,443,379	2.8
Tanga	771,060	1,037,767	2.7

Table (ctd...)

Region	Population 1967	Population 1978	Annual growth rate 1967/78 (%)
Iringa	689,905	925,044	2.7
Singida	457,938	613,949	2.7
Mara	544,125	723,827	2.6
Lindi	419,853	527,624	2.1
Mtwara	621,293	771,818	2.0
Coast	428,041	516,586	1.7

URBAN POPULATION GROWTH

Urban population of the Mainland has been growing at a very high rate since 1967. This is essentially true even if one adopts a more rigid definition than the one used in the 1978 national census. This rate of growth, as indicated in Table 40, cannot be expected to operate indefinitely in the future. However, on the other hand, it would be unrealistic to expect any dramatic decline in this rate in the near future. Economic and other measures necessary for such decline would take time before they could register appreciable impact.

It appears that district and regional headquarters would continue to absorb most of the growth of urban population, though it is likely that other localities would grow in their number and, to a less marked extent, in their contribution to the overall urban population.

Table 40 TOTAL POPULATION, URBAN POPULATION, PERCENT URBAN, aAND RATE OF GROWTH OF URBAN POPULATION FOR THE MAINLAND, ZANZIBAR AND TANZANIA

Year (census)	Total Population	Urban Population	Percent Urban	Urban Annual growth rates (intercensal)
Mainland				
1948	7480400	197300	2.64	
1957	8788500	364100	4.14	6.81
1967	11958654	685092	5.73	6.32
1978	17036499	2257921	13.25	10.84
Zanzibar				
1948	264200	52700	19.95	3.47
1957	295600	72000	24.36	3.43
1967	354400	101475	28.63	3.85
1978	476111	154979	32.55	
Tanzania				
1967	12313054	786567	6.39	
1978	17512610	2412900	13.78	10.19

The level of urbanisation in Zanzibar was higher than that in the Mainland, but a striking feature of urbanisation of Zanzibar is the relatively persistent slow growth of urban population. Also the near constancy of these rates at low level indicates that rural-urban migration has played only a secondary role in the growth of urban population with most of the growth attributable to the natural increase. This, once again, sets Zanzibar at variance with the Mainland where rural-urban migration appears to have played a more significant role. General economic conditions coupled with size and density of population, level of urbanisation already attained, spatial differentials in economic opportunity, relative ease of accessibility between different parts, political and other links with the Middle East and other countries of East Africa, all tend to indicate that no dramatic upsurge in urban growth, at least in the near future, is likely to occur in Zanzibar.

GROWTH OF SEGMENTS OF POPULATION

For detailed analysis, it is sometimes desirable to compute growth rates of sub-populations or special groups of a population. For example, the growth rate of females and males separately or the growth rate of a population in a certain age group (Active population, children under 15 years etc.) may be needed. Sometimes it is also important to look at the growth rate of the population in various sectors of the economy - for example, the population in Agriculture, Industry etc. In any case all these computations are not a problem, if data is available and comparable between two points of time. Table 41 gives some of these computations which also serve as socio-economic and demographic indicators.

Table 41 SELECTED SOCIO-ECONOMIC AND DEMOGRAPHIC INDICATORS FOR THE MAINLAND, 1978: RURAL AND URBAN

Indicator	Urban	Rural
Sex ratio	108.4	94.4
% under 15 years	40.4	46.9
% 60 years and over	3.4	6.5
% Single, males (12 years and over)	45.5	40.8
% Single, Females (12 years and over)	28.6	24.5
Child-woman ratio	701	821
Dependency ration (%)	77.9	114.6
% Literate, males (10 years and over)	84.4	61.5
% Literate, Females (10 years and over)	59.8	35.6
% Literate, total (10 years and over)	73.0	48.1
% with no formal education (10 years and over)	27.7	51.7
% with primary school completed (10 years and over)	21.6	6.4
% with secondary school completed (10 years and over)	0.36	0.01

11.2 POPULATION STRUCTURE

In addition to knowledge of the growth of a population, we also need to know its sex and age structure because these two important characteristics of a population have considerable impact on demographic, socio-economic behaviour.

Many types of planning such as planning of community institutions and services like health services, manpower development policies etc. need data by sex. Social and economic relationships are influenced by sex composition of a population. So also are social roles and cultural patterns. Demographic variables like fertility, mortality and migration have already been noted to be influenced by the sex composition of a population.

For evaluation of statistical data, information by sex provides an useful tool.

Age information is yet another important aspect of a population because the incidences of demographic, socio-economic variables are influenced by age and in combination with sex is an indispensable piece of information not only for planning and policy making but also for evaluation of data.

Age and sex structure of a population are influenced by the demographic factors of fertility, mortality and migration. Sex ratio at birth is another important factor affecting the sex composition of a population. Sex differentials in mortality and migration bring in variations in the sex composition of populations. Coverage and quality of enumeration are also responsible for observed patterns of age-sex distributions of populations.

Generally, 'young' populations and populations with high birth rates tend to have higher overall sex ratio than 'old' populations and populations with low birth rates because of the excess of boys among births and children and the deficit of men among older persons (higher male than female mortality). A population is defined as 'young' if its median age is less than 20 years, as 'old' if its median age is more than 30 years and 'intermediate' if median is between 20 and 29. The median age of Tanzania being 15.7 years is thus a very young population.

Another criteria for considering a population as 'old' may be based on proportions of older persons. In developing countries, generally there will be less aged persons than in developed countries with high levels of expectation of life at birth. In addition to mortality, the high level of fertility usually observed in developing countries results in a relatively lower proportion of old people. As mortality is reduced, the population tends to an 'old' one and this process gets accelerated if fertility also is reduced. The problems of 'young' and 'old' populations are different in terms of socio-economic needs and dependency ratios (child and old age) would be different.

.../146

12. POPULATION COMPOSITION - EDUCATIONAL AND ECONOMICALLY ACTIVE

12.1 EDUCATION

Education statistics collected in censuses and surveys are normally that on literacy and formal education. It is difficult to define and collect statistics on informal education. Education is an important variable which cannot be ignored in demographic analysis.

In the Tanzania's census of 1967 and 1978 it was sought to determine literacy rates by including in the census questionnaires the question "can you read and write in Kiswahili?"

The following table presents the situation in 1978.

Table 42 PERCENT LITERATE BY SEX - MAINLAND AND ZANZIBAR

AGE	MAINLAND		ZANZIBAR	
	M	F	M	F
10-14	76.2	72.5	83.1	73.6
15-19	83.5	60.0	78.9	56.4
20-24	76.9	47.4	73.3	43.7
25-29	74.5	38.6	65.9	31.4
30-34	70.8	28.3	56.1	18.7
35-39	63.8	21.5	50.4	16.1
40-44	55.2	15.3	40.0	8.8
45-49	48.9	12.2	37.2	7.9
50 +	32.9	5.6	22.0	3.1

On education, the question asked was 'What is your highest formal school education?' and for those still attending school the 'class' they are attending was noted down whereas for those who have completed schooling, the information collected pertained to highest 'class' completed.

From these information, it is possible to calculate attendance rate and highest level attained for those who have completed schooling. For example, whereas in Mainland 42.5% of those aged 15-19 were attending school, in Zanzibar it was 40.3%. Similarly those never attended aged 15-19 years constituted 27.9% in Mainland and 33.1% in Zanzibar and the balance 29.6% in Mainland and 26.6% in Zanzibar were returned as 'ever attended'.

Again for those aged 10 years and above who have ever attended school whereas 92.8% in Mainland had completed standards 1-8 it was 68.3% in Zanzibar. For standard 9-14 on the other hand, Zanzibar had 38.1% and Mainland only 3.3%. Both had .4% having completed university or higher education. Thus it is clear that among those who go in for education in Zanzibar, there is a high proportion who continue for post primary education.

.../147

12.2 ECONOMIC ACTIVITY

Definitions

All persons consume goods and services but only a part of the entire population of a country is engaged in producing such goods and services. Those people who engage in the production of goods and services are said to be economically active. Those who are not engaged in the production of goods and services are economically inactive and are said to be dependant. Economically inactive population include mainly young children, students, the aged, the disabled, underemployed home-makers (mainly housewives) and all others who are not working.

Labour Force consists of all employed persons (economically active population) and those unemployed persons who are able to work and are seeking work for pay or profit.

Measures

There exist various measures for economically active population depending on the type of data available. These measures are generally referred to as economic activity rates or labour force participation rates. These rates can be calculated on the basis of the whole population or on specific ages and separately for males and females. The ages 15 to 64 years is usually regarded as the working age. Therefore if the exact size of active population is not known, those aged 15 to 64 years are usually taken as an estimate for active population. The questions on economic characteristics of the population in the 1978 census were 'what are you usually doing to earn your living' and 'what is your economic status'. Tabulations pertained to the work force by age classified by those with wage and salary employment (permanent, temporary and casual) and others like own account and family workers. In addition tables have been prepared of population 5 years and over by main occupation.

For example, in Zanzibar among females at age 25-29, there were 9116 in the work force out of a total of 16017 females giving a participation rate of 56.9%. Out of this, 7027 or 77.1% had agriculture as a main occupation.

13. SIMPLE PROJECTIONS AND ESTIMATES

Introduction

The study of population growth has one specific practical application - to calculate population trends for the future. The product of such calculation is given various names including future population, forecasts, extrapolations, estimates or projections. In either case, they all imply extending some plausible pattern of growth from the past into the future and vice versa.

Population projection is simply a term intended to imply no more than an illustrative calculation based upon certain given assumptions.

At the present time, the most complete and reliable source of information on the population of countries and their geographic subdivisions is the census based on a house to house enumeration. Population is constantly changing, sometimes quite rapidly. Hence census statistics which are normally collected once every ten years are not adequate for most current needs. Planners, researchers, educators, politicians etc. need current data (in each year) to plan and evaluate programmes.

In order to meet the need for basic population figures more fully, a wide variety of estimating techniques, including the use of sample surveys, has been developed. Like the census, sample surveys are too expensive to conduct each year and cannot provide data for past or future dates. Nonsurvey or analytic techniques involving the use of vital statistics data, immigration data, other data symbolic of population change, and mathematical methods however are relatively inexpensive to apply and can be used to prepare estimates for past and future dates as well as for current dates.

Estimates fall into three broad groups. These are:

(1) Intercensal estimates, which relate to a date intermediate to two censuses and take the results of these censuses into account.

(2) Postcensal estimates which relate to a past or current date following a census and take that census and possibly earlier censuses into account, and

(3) Projections which relate to dates following the last census usually future dates, for which no current reports are available. Both postcensal and projections can be regarded as extrapolations, and intercensal estimates as interpolations taking these terms in their more general sense.

Estimates may be for the total population of a country or geographic subdivisions of a country or classes of areas within the country (e.g. urban and rural areas) or a particular class of population distinguished by age, sex, race etc.

The framework of projections

No war or natural disaster and no economic fluctuations of a cyclical nature;

Choice of the projection technique and Determination of base period estimates;

Projections of growth components is commonly founded on some version of the theory of demographic transition: Sometimes projections are modelled on the experience of a particular country more advanced in the demographic transition than the one for which a projection is being made; in other cases assumptions as to the

forms of the future transitions and trends of mortality and fertility are drawn according to the intuitive judgement of the authors.

What is required is respect for the specificity of each epoch and each population.

Among other relatively simple ways of projecting the growth components (fertility, mortality and migration) are (a) assumption that they will remain constant at the base period levels (b) extrapolation of past trends and (c) projections to set targets assumed to be reached at specified future dates. Extrapolation or the assumption of no change is often used when no satisfactory basis for other assumptions is found while target setting is often practiced in projections of migration commonly assuming that net immigration or emigration will decrease to the vanishing point. As far as possible, assumptions as to the future of the demographic variables should be formulated within a conceptional framework of their relations with expected social and economic changes.

Projections can be obtained for one area on the basis of changes in some other similar area or more inclusive area for which a projection is already available. This includes

(i) Distributing the projected population of an area among its subdivisions taking account of the past proportionate distribution (ratio methods) and (ii) Projecting the growth components for an area on the basis of changes in these components of a similar area.

Usually four varients - constant, low, high and medium are prepared varying the parameters suitably allowing some flexibility to users in terms of varied needs and applications.

Length of projections depends on type of area, the needs to be served, the conceptions of the problem by the analyst and the available resources in general.

Long term projections (25 years) are needed in connection with water and forestry resources, transportation and recreational facilities, planning for food production, etc. Middle range projections (10 - 25 years) are required for planning educational and medical facilities and services, housing needs. Short term projections (10 years) are needed for short term economic analysis. Simple methods may be used for short/middle range projections; more elaborate methods are used for the longer term projections.

Frequency of projections is a function of:

the extent to which current projections are out of tune with current estimates, the mere passage of time, availability of new data, advances in new data and advances in new methodology.

.../1'

It is important to observe that:

(a) More accurate estimates can generally be made for an entire country than for the geographical subdivisions of the country. The national population is much more likely to be a closed population than is that of a subdivision, of the country. Moreover, when there is immigration, it is likely to be registered for administrative reasons while internal migration will go unrecorded. In general, more direct data, data of better quality and more information on how to adjust these data for deficiencies, are available for the larger areas, particularly for entire countries, than for the smaller areas. Similarly more accurate estimates can generally be made for the total population in an area than for groups based on demographic characteristics of the population of the area.

(b) The accuracy of the estimate will greatly depend on the accuracy (quality) of the base data used and the correctness of the estimation method and assumptions made. The accuracy of the final estimate will also depend on the length of the estimation period. Too long term projections are likely to be unreliable.

Essentially there are 2 approaches - mathematical curves and models and cohort - component method. Recently, less and less use is being made of mathematical methods although they are not abandoned. In specific cases, use of mathematical methods like in subnational and sectoral projections may be the only avenue available.

13.1 MATHEMATICAL CURVES

One of the easiest methods of population estimation especially projections is the use of mathematical curves. If a mathematical growth formula (curve) is assumed and the base population and rate of growth are known, then the population estimate for any year can be obtained from the formula by simple calculations. The main problem is to get a mathematical curve which would adequately fit with the growth of a given population especially over long periods of time. As a result this method is very unreliable for projection purposes and is normally used for short-term projections only.

Linear growth

Linear growth implies that the population is increasing by the same quantity (number of persons) each period (year). This however rarely happens in any real situation. Most populations are growing in such a way that the number of people added to them increases each year. Therefore in almost all cases population growth can only be presented by (an upward) curve and not a straight line. Therefore linear growth formula has little application, excepting where migration of a fixed number of persons per year is operating and submerges natural growth.

Geometric growth

Perhaps the most commonly used growth curve in population projections is the geometric formula $P_t = P_o(1+ r)^t$. Geometric curve assumes a growth by constant annual compounding and it is sometimes referred to as a "compound interest" curve. Geometric curve is quite suitable for estimating (or projecting) population over short periods. It has some basic weaknesses in that it assumes that the population grow, by compounding after a certain period (year) whereas in actual fact population growth is continuous.

Exponential curve

Since population grows continuously, the exponential curve which give continuous compounding is sometimes used. Exponential curve is of the form $P_t = P_o e^{rt}$.

Other growth curves

Other growth curves like the modified exponential, the Gompertz and logistic have also been used in population projections.

Projected populations using growth curves are obtained by substituting the value of 't' the time period in the mathematical equation derived based on past data. For example, from the equations fitted in Section 7.4 we can note that if we put t = 1978 corresponding to x = 21 we will get the projected population of 1978.

From the linear equation

$$y = 9633263 + 241966x \text{ we get}$$

$$\text{Pop. } 1978 = 9633263 + 241966(21) = 14714549$$

and from the exponential equation

$$\ln y = 16.062 + 21(.025) = 16.587$$

$$\text{or} \quad y = 15982421$$

These figures can be compared with the enumerated population of 17512611. Whereas the linear growth underestimated the population by 19% the exponential growth formula was off by about 10% only. The past growth has been slow and the curves depict this. The accelerated growth of 1967-78 was not anticipated by either of the curves.

.../152

DEM-100
POO

13.2 COMPONENT (COHORT) PROJECTIONS

In these, separate projections of the components of population change are made i.e. fertility, mortality and migration are projected and in conjuction with the base age-sex structure, these are used to carry forward the population. Since the base age-sex data may be affected by errors, biases and deficiencies, smoothing and adjustment for coverage and content are first done and then age specific fertility, mortality and migration rates are used to project the population.

Since the estimation of births in the projection process requires the average number of females by age groups in the reproductive span, usually the operation is carried out with respect to mortality and migration and the average is usually taken as the arithmetic mean of the base and projected values. (Due to availability of data, it is conventional to carry out the projection at intervals of five years and in five years age groups. However, if single year or ten year of age and other parameters are available, then such projections may also be possible).

For projecting the survivors of the base period, the use of age specific survival ratios from life tables would be necessary (by sex). Thus with five year of age group data and five year survival ratios, we carry forward the population by five years and the population would be five years older. Hence children aged 0-4 years (who are the survivors of the births of the 5 year period) need to be estimated by first obtaining the births (by multiplying the average female population by age specific fertility rates or using SAABR or synthetic fertility schedules (model patterns) with given TFR). The sex proportion at birth can then be used to obtain male and female births. Symbolically, if P_x^t is the population aged x at time t and S_x^t is the survival ratio for the population aged x for the one year since time t, then population aged x+1 at time t+1 is

$$P_{x+1}^{t+1} = P_x^t \; S_x^t$$

assuming that migration is nil. Otherwise, an appropriate figure of migrants during the period at that age can be added to obtain the projected population one year older at a time one year later. Similarly for five year grouped data.

To obtain the number of births in a five year period the formula is:

Total no. of births = $\frac{5 \text{ SAABR}}{1000}$ (weighted sum of average of female population aged 15-19, 20-24, ... 40-44 with respective weights 1, 7, 7, 6, 4, 1).

(if SABBR is used) and

= 5(Sum of product of average female population aged 15-19, 20-24, ..., 45-49 and corresponding age specific fertility rates).

(if age specific fertility rates are used)

The number of male and female births would be:

Male births = Total births (Male proportion at birth)

Female births = Total births - Male births

The births would then be survived by appropriate survival ratios and be aged 0-4 years. Thus we obtain the projected population aged 0-4 years.

Let us illustrate the method with the projection of the female population of Tanzania for the period 1980-85. (here mid 1980 is taken because comparable projections are usually for mid year periods and for years ending in 0 or 5).

Since the census was on 26 August 1978, we carry forward the population for a period of 1.847 years. Since the intercensal growth rate was 3.2% we use the same figure with exponential formula and get the female population as:

$$P_{1980 \text{ mid year}} = P_{1978 \text{ census}}\, e^{(1.847)(0.32)}$$

$$= 8925525\ (1.0609) = 9468961$$

Smoothed age distribution has been taken from the census analysis results. Mortality level has been estimated at more than 11.3 in 1972-73 and hence in 1980-85 it is taken to be 13 and survival ratios are borrowed from Coale-Demeny North model life tables. With TFR of 6.86 estimated and patternoof fertility as shown by census,births have been obtained, split into males and females and survived using appropriate survival ratios. Table43shows the results.

Table 43 Projection of female population,Tanzania 1980-85

	Smoothed % (1)	Population 1980 (2)	Survival ratio (level 13)*	Population 1985** (4)	Average 1980-85 (5)	ASFR (6)	Annual Births (5)(6)=(7)
Births			.865	241485			
0 - 4	18.2	1723351	.929	2088845			
5 - 9	14.4	1363530	.971	1600993			
10 -14	12.7	1202558	.979	1323988			
15 -19	9.1	861675	.976	1177304	1019489	.144	146806
20 -24	8.7	823800	.971	840995	832397	.328	273026
25 -29	7.0	662827	.967	799910	731368	.315	230381
30 -34	6.8	643889	.962	640954	642421	.261	167672
35 -39	5.0	473448	.957	619421	546434	.191	104369
40 -44	4.3	407165	.952	453090	430127	.094	40432

../154

Table (ctd...)

	Smoothed % (1)	Population 1980 (2)	Survival ratio (level 13*)	Population 1985** (4)	Average 1980-85 (5)	ASFR (6)	Annual Births (5)(6)=(7)
45 -49	2.9	274600	.943	387621	331111	.039	12913
50 -54	2.8	265131	.926	258948			
55 -59	2.4	227255	.897	2455711			
60+	5.7	539732	.694	578421			
All ages		9468961		11016001		6.86***	

* North model of Coale-Demeny

** The values in column (4) are obtained by multiplying values in a row in column (2) and (3) and bringing the values to the next row.

*** This is the value of TFR assumed

13.3 SIMPLE PROJECTIONS OF RURAL-URBAN AND SUBNATIONAL POPULATIONS

Projection of urban-rural population

There are several methods for projecting the urban and rural populations - some of them simple but a few very elaborate and sophisticated. We shall illustrate only one simple method of projecting the urban population which utilizes an available national total population projection. This method can be extended to derive age-sex distribution of urban population. As a difference, we derive the rural population.

The method used is mathematical curve fitting and since the progress of urbanization in developing countries has been noted to follow the logistic growth curve, the projection is done through this curve for the calculation of future proportion of urban population.

If proportion urban, on comparative basis, is available for an area at two given points of time, the method has been simplified as follows instead of going through the fitting of the curve.

First, from Table 44, the table years corresponding to the percentage urban at the two time period are noted and the per calendar year change in table years calculated. For a required period of time, first the numbers of calendar years since the last available period is converted to number of table years and added to the table year corresponding to the last period. This gives the table year corresponding to the required time from which using the table, the percentage urban can be read off. This method is the UN urban rural growth difference method described in detail in this Manual VIII.

For illustration, we have seen in section 11.1 that .39% and 13.78% of the population of Tanzania were reported as urban. These values correspond to table years -269 and -183 respectively (by linear interpolation). Thus a 11 calendar year period corresponds to 86 table years. If we wish to have the percentage of the population urban in 1985 which is 7 calendar years from 1978, the number of table years will be (86 x 7)/11 = 54.7 or 55 table years.

Hence in 1985 the corresponding table year will be -183+55 = -128 For table years -128 the percentage urban is 21.77 from the table. 44

Table 44 Values from logistic curve: $100\, e^{dt}/(1+e^{dt})$ for values of t = table year

Table Year	%	Table Year	%	Table Year	%
-350	2.93	-230	9.11	-110	24.97
-340	3.23	-220	9.98	-100	26.85
-330	3.56	-210	10.91	- 90	28.90
-320	3.92	-200	11.92	- 80	31.00
-310	4.31	-190	13.01	- 70	33.18
-300	4.74	-180	14.18	- 60	35.43
-290	5.21	-170	15.45	- 50	37.75
-280	5.73	-160	16.80	- 40	40.13
-270	6.30	-150	18.24	- 30	42.56
-260	6.91	-140	19.78	- 20	45.02
-250	7.58	-130	21.42	- 10	47.50
-240	8.32	-120	23.15	0	50.00

For obtaining the % for positive values of table years, we use the fact that the curve is symmetric about origin 0. For example the value of the % for table year 80 will be 100-31.00 = 69.00 where 31.00 is the value for table year -80.

The urban population is obtained by multiplying the urban percentage by the projected total population. Assuming for the time being that percentage urban is similar for both sexes, we can estimate the female urban population in Tanzania 1985 using the projected female population in section 13.2 as:

(Projected female population in 1985)(Projected % urban in 1985)/100

= .2177 (11016251) = 2398238

Projection of subnational populations

For planning and policy implementation, in addition to projections of populations of a nation and by urban and rural categories, it is useful

to have projections of administrative areas like regions, provinces, districts etc.

Here also mathematical methods like the ratio, apportionment and growth rate techniques are used in addition to detailed cohort component projections.

13.4 SIMPLE PROJECTIONS OF SCHOOL POPULATION, LABOUR FORCE AND HOUSEHOLDS

School Participation and School population

Projections of the number of children who will be enrolled in school are needed to formulate educational policies and plan educational programmes and, especially, to plan for needed schools, classrooms, and teachers.

Where education is universal, such as Primary education in Tanzania, enrollment depends on the number of school going children in the population. As the population grows, this group will also grow and its projection is determined from the projection of the whole population (by age and sex). Where education is not universal, the size of school enrollment is dependent on population size and structure only through participation rates. Future expansion depends on the policies of the countries and extent of implementation of educational plans.

There are two main methods of projecting school enrollments:

Age-specific Enrollment Rate Method:

The most widely used procedures of preparing projections of school population has been to employ assumed age-specific or age-sex specific enrollment rates in combination with the projected population by age and sex. The assumption relating to enrollment rates may be quite simple. One way would be to assume the continuation of the current or recent enrollment rates. Another way would be to extrapolate (project) past trends of enrollment rates. In this case the past trends may be assumed to continue as observed or projected taking into account changes in other socio-economic variables. Alternatively, one can borrow rates prevailing in more advanced countries whose conditions are assumed similar to the future conditions of the country under consideration. Where terminal rates are set, rates for intermediate years are usually obtained by some form of mathematical interpolation.

When projections of school enrollment are made by the enrollment-rate method, they rarely contain detail on grade or school level. To obtain such figures, one procedure is to prepare projections by age by the enrollment-rate method first, and then distribute the projected total enrollment at each age by grade or school level on the basis of recent census, survey, or administrative data (holding the distribution constant or extrapolating it).

To illustrate the projection, let us consider the enrollment rate at age 15-19 in 1978 which was around 42%. Assuming that in 25 years

i.e. by 2003 AD, the participation will reach 60% we can interpolate the rate for intermediate periods. For 1985 for which we have a population projection (for females) let us estimate the female school population. Linear interpolation of the rate will produce a value of 47%.

Using this percentage in conjuction with the projected population of that age group will give the projected school population. Assuming that female participation is not much different from total participation, we can obtain the female school population as:

= (Participation rate at age 15-19) (Projected population 15-19)
= .47 (1177304) = 553333

Cohort Method:

Projection of age-specific enrollment rates on a period basis in the framework of an enrollment rate method has been the predominant method of making projections of school enrollment. For the less developed countries at all school levels, the possible variation in enrollment rates renders projection of enrollment rates a task fraught with great uncertainty. Other methods are therefore employed. Cohort method is used to secure enrollment projection by grade. In this procedure the number of enrolled persons by grade is carried forward to each subsequent calendar year by use of projected grade-retention rates or grade progression rates, representing the proportion of children in a given grade who will advance to the next grade in the course of a year. A historical series of grade-retention rates may be developed on the basis of survey data or data from administrative records of the school system and then projected forward on an annual basis. Projections of the new first grade cohorts may be derived from population projections for the "entering" ages on the basis of the age-grade relationships shown in the census.

Economic Activity Rates and Economically active population

Economic activity rates are projected in order to estimate the future size and structure of the working population. Projections of economically active population are needed to give an indication of the number and characteristics of the workers who will be available for employment in future years so that appropriate plans and policies can be made. The economic activity rates are projected based on assumptions like:

a) The trend in the activity rates of the economically active population for future years will be an extrapolation of the past trend; b) Current activity rates will be maintained in future years; c) Activity rates in future years will be the same as those of other more advanced countries or the same as current rates in the more developed areas of the country concerned; or d) Activity rates will depend on projected changes in such factors as the economy's manpower needs, school enrollment, urbanization etc.

.../158

Any estimate of the size of economically active population based on any of the assumptions will become unreliable if the assumption prove to be unfounded. Quite often the activity rates are projected using two or more of the assumptions cited above. The choice of method depends on the statistics available and on judgements regarding the prospects for economic and social development in the country concerned. Incountries where non-availability of statistical data makes it impossible to determine past trends or even, in some cases, to ascertain the current situation, future activity rates are often calculated on the basis of examples drawn from other countries which in the past followed the same course as the country concerned. If extrapolation method is used, future economic activity rates can be computed by linear extrapolation of observed trends. This method is the simplest and the most frequently employed but it is unsatisfactory for a number of reasons the most important being that it can yield percentages above the maximum of 100. A geometric projection has a similar defect if the trend being projected is a rising one. Use of logistic curve or other curves with an upper limit of 100% will avoid such anomalies.

Projection of Households

A household is a socio-economic unit consisting of individuals who live together and make common provision for food and other essentials for living. Households usually occupy the whole, part of or more than one housing unit, but they may also be found in camps, in boarding houses or hotels. The head of the household is defined as the person who is acknowledged as such by the other household members. Usualy this is the person who bears the chief responsibility for the economic maintenance of the household.

Projections of households are required for many uses, particularly those which depend on information regarding future numbers of consumer units. A principal use of projections of households is in deriving projections of housing needs, but they are also needed to anticipate demand for other consumer goods and services required by households as a unit.

Many methods for projecting households have been devised and applied by different countries, institutions and individuals. The methods used vary widely depending upon the type of data available and the needs served by the projections. Among them are:

Simple Households-to-Population Ratio Method (Average household Size)

Because of insufficient data on households heads, it is not possible for developing countries to make an elaborate projection of households which take into account the various factors affecting their future growth and structural changes. In these countries, it is frequently necessary to resort to population census data for estimating the future rate of growth of households. A crude estimate is obtained by taking the rate of growth of total households to be equal to the rate of growth of the population or equivalently, by taking the same ratio of the number of

households to the total population and applying it to the future population projections already prepared. A better estimate of the future member of households may be obtained by applying the ratio between the number of households at the base year and the adult population to the projections, since household formation is usually confined to this section of the population. In many countries, however, the number of households may grow at a considerably different rate from that of the total or adult population. A growth rate different from that of the population may therefore be used for households.

Headship Rate Method

This method employs available population projections by sex and age (and sometimes marital status) as its base. The population is therefore classified by sex and age and if possible by marital status. The projected number of households is obtained by adding up overall age groups the product of population and headship rate.

The main methodological problem in the headship rate method of projections is how to estimate accurately future levels of headship rates specific for sex and age or for sex, age and marital status. Three methods based on some basic assumptions about the future trends of the rate which produce easy projections are:

a) Constant rate method; b) Extrapolative method by using annual average change of rates in the past or by applying a simple Mathematical formula on the basis of past trends; and c) Normative approach drawn up according to the country's social and economic development programmes.

14. IMPLICATION OF POPULATION GROWTH

14.1 GROWTH RATE, AGE-SEX STRUCTURE AND DENSITY

As a result of changes in fertility, mortality and migration rates and consequent to age-sex composition differences, populations grow or decline at varying rates. The rate of growth of a population is thus a determinant and a consequence of future population size, structure and composition.

A rapidly growing population imposes pressures on the various aspects of an economy and society and in situations where the present position is such that some of the basic needs of the population has not been fully met, the implication fof the future is obvious.

For example, the projection of the female population presented in section 13.2 clearly shows that from a base 3.2% growth rate, the rate has increased to 3.3% even inspite of an assumption of fall in fertility. Had fertility not decreased, then the growth rate would have been even higher. Again, if mortality fell faster than the assumed normal pace, then also the growth rate would have been higher and vice versa. The age structure which is a determinant of fertility-mortality is also a consequence of the interplay of these vital factors. Similar is the case of sex composition.

Age-sex structure

When mortality falls with no corresponding fall in fertility, the rate of growth of the population will accelerate and since mortality falls generally more among children than among adults, the growth of the child population can be much in excess of the other segments. When fertility falls rather rapidly, then the population becomes aged. Thus the age structure is determined by the levels, patterns and trends of fertility and mortality. On the other hand, the vital parameters themselves are determined by the age structure of a population as we have noted that the incidence of these vital events vary over the age range.

The sex composition also is a cause and consequence of the interplay between the vital parameters in addition to the level of the sex ratio at birth. For example, one observed phenomenon in circumstances of falling mortality is the decline of the sex ratio of the population because of the female over male advantage in survival.

From the projection in section 13.2 we see that the child population in 1985 has grown at 3.8% as against the 3.3% noted for the total population and has become 19% of the total population compared with its share of only 18.2% in 1980.

The implication of varying growth rates of the various age-sex segments to the economy and society is well known.

Density

As the population of an area increases, so does the density of population. But it may be argued that the present density is not very high and hence an increase may not matter much. But what is forgotten is that the increase in density may not occur uniformly and the already over crowded areas like towns and cities may be the ones bearing the brunt of the increase.

In Tanzania admittedly the density of population in 1978 was low and only less than 20 and with a 3.3% growth between 1978 and 1985, the density would still be only around 25. However, the urban growth of more than 10% indicated in section 13.3 clearly brings forward the fact that the problem of population distribution may be compounded by the continuing high growth rate of the population. As a matter of fact, if the projections depict the scenario expected in the short period of 7 years since 1978, then the rural population will grow only at 1.5% and the problem of food production will emerge.

14.2 EDUCATION, LABOUR FORCE, AND DEPENDENCY

Interest in population growth goes beyond aggregate changes in total population of a country over time to include changes in magnitudes of identifiable groups which compose that population. This is due to the fact that implications of growth of a country's population

manifest themselves in functional or target groups. We are used to dividing the population of a country by sex and age groups. The subdivision of the population can be done by single year groups or by larger age groups depending on the purpose of analysis and use of the population statistics. For example since the roles of individuals in society are normally performed according to age, the population can be divided into children (under 15 years of age), Adults (15-64 years) and old people aged above 64 years or into labour force and dependent population etc. For policy formulation and development planning purposes it is important to trace and forecast accurately the changes in magnitude (absolute and relative) of the various population groups. Changes in each group in most cases have different action implications to the state.

Our interest here is to examine the impact of population growth on education, labour force, dependency ratio and housing. This involves dealing with particular age groups classified as school age population, working age population, dependent population and family formation age etc. Population projections provide estimates of the relevant segments.

Education

Education is a very important factor in the socio-economic development of a country. It constitutes an essential pre-condition to economic growth which among other factors depends on human resources. The development of human resources involves imparting, through education, skills, knowledge and attitudes conductive to economic change and progress. This attribute of education makes it necessary to integrate educational programmes into the whole planning process. One of the main objectives of development and social well-being is to provide the individual an opportunity to develop his potentialities to the highest degree possible. With respect to this aim it is believed that there is a certain level and content of education which opens the scope of the individual's development.

The provision of education in Tanzania aims at creating the necessary manpower for economic growth and at providing the individual with the tool for the realization of his development potential. However, emphasis on the two aims seems to differ with time. At independence, manpower development was of prime importance in order to man the vacancies left by the departing colonial servants and for localization purposes. The Mainland's first long term plan had self-sufficiency in skilled local manpower as one of the main objectives. The problem of shortage of skilled manpower is still intense in the country. However the nature of the problem has been-changing over the period of the plan. Acute shortage at high level of skilled manpower is now being experienced and the need to develop the right attitude conducive to economic progress is manifesting itself clearly with the passage of time. Due to this change in the nature of the education problem we have seen a growing emphasis on primary education towards the end of the plan culminating on working for Universal Primary Education by 1980, as opposed to priority given to the expansion of higher education for manpower development purposes.

Primary Education

Basic education in Tanzania is provided in primary schools (for the Mainland) and plus 3 years of secondary education for Zanzibar. This does not imply that these are the identified levels of education required to open the mind of every citizen to the development of potential.

Basic education is a right to every citizen in order to give each individual an equal opportunity to advance. This fact compelled the state to institute measures to ensure basic education for all at the earliest time possible. For the Mainland the aim was to attain universal primary education by 1975. Universal Primary Education means basic education to all children who reach primary school age, 7 to 14 years. Persons in this age group must enrol and follow classes until they complete primary school education. In order to develop, maintain, and improve universal primary education a close monitoring of the growth overtime of the school going age population is important, because of the implications of population growth for teacher training, school construction and the provision of other facilities. The growth of the primary school going age group between 1978 and 1985 in Tanzania can be obtained by comparing the size of the child population aged 7-14 years at the two periods.

We shall illustrate the idea using the female population of 1978 and the projected female population of 1985. Since single year values are subject to errors and any way the projection gives only figures by five year age groups, we obtain the population aged 7-14 years by first interpolating the population aged 7-9 years from the group aged 5-9 years by taking 60%(3 out of 5)ages of the value and adding it to those aged 10-14 years.

With this approximation, from Table 6 we obtain the female population aged 7-9 years in 1978 as .6(1412285) = 847371

Thus female population aged 7-14 years = 847371 + 1034802 = 1882173

Similarly from the projected figures we obtain

female population aged 7-9 years = .6(1600998) = 960599
and female population aged 7-14 years = 960599 + 1323988 = 2284587

This implies a 3.9% annual growth rate which is much higher than the 3.3% annual growth rate indicated by the total population. In other words, the child population is growing faster than the total population.

Since participation rate also is growing, the school population, which is a product of two increasing variables' like population and its participation, will be growing much faster than 4%. This implies need for increased investment, manpower etc. or in other words diverting scarce human and material resources for present consumption from investment needs.

Labour force

The growth of the labour force is determined by the growth of

population in the relevant age group, the sex composition and the changes in participation rates.

Since participation rates at adult ages do not generally fluctuate over time or age range, it is the growth in the size of the relevant age group which determines the size and growth of the labour force.

Assuming that persons aged 15-59 are mostly in the labour force we can obtain an idea of the size and growth of the labour force by comparing the group over time. For example, in 1980 there are 4639790 and in 1985 it is 5423754 i.e an annual growth rate of 3.1%. This is lower than the rate of growth of the population. In other words, the economically active are not growing as fast as the total population even when we take the participation rate same for both periods. If participation rate falls due to increased numbers going for education etc. the gap would be even wider.

Dependency

The child population aged 0-14 years grew from 4289439 in 1980 to 5592252 in 1985 i.e. with an annual growth rate of 5.3% about double the growth rate of the population. The old age population i.e. those aged 60 and above did not increase much but this may be due to problems in adjustment of the age distribution of the base population.

The dependency ratio ie. ratio of (children aged 0-14 years + old persons 60 and above) to those aged 15-59 in 1980 was 1.04 and 1.03 in 1985 ie. there was an apparent fall. This is due to the slow growth shown by the old age population perhaps consequent on errors in adjustment to the old age segment in the base period. However, the child dependency ratio remains high at 92.4% at both periods even inspite of the assumed slight fall in fertility in the projections. This is due to the fall in mortality which affects the child population much more than the other ages.

14.3 HOUSING, FOOD AND HEALTH

Housing

We have already seen that housing needs are directly related to the size of the adult population through the headship rate. Thus an increase in the size of this segment automatically would imply increased need for house construction unless one wishes a deterioration in the housing standard.

Migration and urbanization are other factors influencing housing needs and we have seen that urbanization is increasing rapidly in the country. Another phenomenon is the formation of more nuclear families resulting in increased demands for housing.

Food

The consumption of food varies with the size and age-sex structure of a population. Thus changes in the population size and its structure

will have repercussions on the need for food. Again the quality of food is another important fact dependent on age structure - children and old persons need better quality food as compared with the predominant starchy - carbohydrate composition of food consumed by adults.

Migration and urbanization not only result in depletion of the rural population (the food producers), the types and varieties of food consumed in towns and cities call for production of such food items as vegetables, fruits and meat instead of bulk food like cassava, maize consumed by rural population.

Also pregant or lactating women need more nutritious food and a shortage can lead to deterioration of health of mother and child. In a society with high fertility, the proportion of such women would be large.

As most of the labour in the agriculture sector is carried out by manual process and the consumption needs of manual work is larger than that for sedentary workers, the quantum of food required will be larger.

Health

The size of a population will determine the health needs only partially because it is its age and sex structure which are important factors influencing health needs and services. Children, old persons, pregnant and lactating women consume a large share of the available health facilities and services.

The environment, nutrition and the type of work also affect health status. The spread of disease through migration is well documented. Education, spread of knowledge and other channels of communication affect health status of a population.

Thus we note that population size, growth, structure, composition, distribution and characteristics (socio-economic) have implications on the various aspects of life.

15. USES OF DEMOGRAPHIC DATA IN PLANNING AND POLICY FORMULATION

Stages in the Planning Process

To the serious planner or policy-maker, questions will occur which can be answered by the statistician. To illustrate this we will first consider the planning process. What is it like? Five states can be distinguished.

.../165

I.	SETTING GOALS/TARGETS	Formulation of <u>what</u> we want, where we want to go
II	DESIGN OF THE PLAN	Formulation of <u>how</u> we will reach target/goals; time schedule; which means are needed? setting subtargets
III.	EVALUATION OF SITUATION	<u>Examination</u> of trends during a recent period of time; is the plan realistic? can targets be met within reasonable time? <u>If no</u>: go back to II (adjust design)
IV.	IMPLEMENTATION OF THE PLAN	<u>Action</u>: go the designed path
V.	PERIODICAL EVALUATION	Were targets/goals reached? <u>If yes</u>: go on; <u>If no</u>: why not? Does plan need adjustment? <u>If yes</u>: go back to II <u>If no</u>: go on

At stages IV and V, questions will occur to the planner which can partly be answered by an adequate statistical information system.

<u>Setting Goals</u>

The following list of aspects which should be considered by planners and politicians can be prepared:

1) Food, nutrition, clean water, sanitation; 2) Shelter; 3) Clothing; 4) Health services; 5) Education; 6) The Social and cultural context; and, 7) The economic context (inflation, national income, balance of payment).

The policy aims now, should be translated into activities with <u>measurable goals</u>. Measurable targets should be set for all aspects 1 to

.../166

7 mentioned above. Only then we will be able to evaluate the success of our policy and the effectiveness of our efforts.

The Role of Demographic Data

Policy goals will often be expressed as changes from the current level (say, we want an annual growth in food production of 5%). However, all planning and policy activities are - or should be - aimed at people. Therefore, we need to know how many people there are, where they live, what their approximate age is etc.

The same holds for stage V of the planning process (evaluation of our efforts). Statistical data on birth, death, food production, consumption, national income etc. assume significance only if compared with population size. That is why figures are calculated per capita; per 1,000; per 100,000. That is why percentages are computed (illiteracy ratio, vaccination coverage ratio's etc.). Similarly, a growth in food production is only real if it is higher than the growth in population. In fact, it is estimated for SubSaharan Africa, that since the 1960's, though absolute food production remained constant or showed minor growth, food production per head of the population decreased by 2 per cent each year (World Bank Reports quoted in African Business and the Daily News). Commonly used - though rough - indicators of a country's (a region's) level of health care are hospital beds and medical staff per 1,000 population, although "The number of water taps per 1,000 persons is a better indication of health than the number of hospital beds" (Halfdan Mahler, Director-General, WHO). Again, levels of housing can be approximated by number of persons per house or per room. In an international context or in time perspective, figures on national income only have significance if calculated per head of the population.

Demographic Changes: Long Term Effects

Example 1

Assume that a decline in fertility is observed. For example, a reduction of 2 children during childbearing age (say from 6 children per woman to 4). The implications for the planning are enormous. There will be more food available for the children who are born (10% of children below age 5 who die in Zanzibar hospitals die from nutritional deficiencies as the first cause of death). School classes can be extended at a lower rate. Schedules for the education of teachers must be reconsidered. The lower number of babies will, after 20 years, result into a lower number of adults who - with fertility still at its reduced level - will produce even fewer children.

Example 2

Assume that fertility remains unchanged, but that infant mortality from communicable diseases is reduced. The latter is very probable as an Extended Programme on Immunization (EPI) is due to be implemented soon (approximately 30% of children below age 5 who die in Zanzibar hospitals die due to measles and 3% due to tetanus). Moreover, a big Malaria control

programme will be started in the course of this year (16% of causes of death for children below age 5). What will be the consequences? An increase in child population and consequent demand on health, education, food and other resources. The future demand for jobs and housing also are obvious.

Example 3

Industrialized countries have seen decreasing fertility and mortality. This development is applauded and is inevitable. However, problems occur. The proportion of aged people in these countries is 10 to 15% now, but is expected to grow to 25% and over by the year 2025. How to finance the old age pensions? Where to find nurses for the old people's homes?

From the above examples it is evident that demographic changes have their impact on a very long term.

In making population projections we will have to make educated guesses about many elements. (For example, if we observed reduced fertility, will it go on to decrease? or remain at the reduced level?). In many countries, developing and as well developed, population projections must be revised again and again due to changes in vital parameters.

Planners want to have things under control, it is their job to control and adjust. So why not control and adjust population size?

Planning Demographic Changes

In many developed countries the reduction in fertility has come more or less by itself. Developing countries, faced with sudden and unexpected population growth, while production of food and other goods did not keep pace, have in many cases been forced to consider an active family planning policy.

Since population size, structure, growth and distribution have implications on the economy and society, plans and programmes should take care of these.